

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À  
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE  
APPLIQUÉES

PAR  
KHALED SALEM

À PROPOS DE LA COVARIANCE LIMITE DU PROCESSUS DE COPULE  
EMPIRIQUE

NOVEMBRE 2015

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

# Remerciements

Je souhaite tout d'abord adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire, ainsi qu'à la réussite de mes deux années universitaires.

Je tiens aussi à remercier sincèrement mon directeur, M. Mhamed Mesfioui, professeur titulaire au Département de mathématiques et informatique de l'Université du Québec à Trois-Rivières, ainsi que mon co-directeur, M. Christian Genest, professeur titulaire au Département de mathématiques et de statistique de l'Université McGill, pour leur disponibilité tout au long de la réalisation de ce mémoire, ainsi que pour leur aide et le temps qu'ils ont bien voulu me consacrer.

J'aimerais en outre remercier mes chers parents, mon frère et toute ma famille pour leur soutien tout au long de mes études.

Mon travail a été financé par des octrois accordés à M. Mesfioui ou à M. Genest par le Conseil de recherches en sciences naturelles et en génie du Canada, le Fonds québécois de la recherche — Nature et technologies, l'Institut canadien des sciences statistiques, ainsi que la Chaire de recherche du Canada en modélisation de la dépendance stochastique.

# Table des matières

Remerciements	ii
Table des matières	iv
Table des figures	v
<b>1 Introduction</b>	<b>1</b>
<b>2 Notions théoriques sur les copules</b>	<b>3</b>
2.1 Généralités sur les copules . . . . .	3
2.1.1 Théorème de Sklar et propriétés . . . . .	4
2.1.2 Utilité pour la modélisation . . . . .	5
2.1.3 Extraction d'une copule . . . . .	6
2.1.4 Copule d'indépendance et bornes de Fréchet–Hoeffding . . . . .	7
2.1.5 Cas multidimensionnel . . . . .	9
2.2 Quelques modèles de copule . . . . .	9
2.3 La famille des copules archimédiennes . . . . .	11
2.4 Simulation de copules bivariées . . . . .	12
<b>3 Processus empiriques</b>	<b>14</b>
3.1 La fonction de répartition empirique univariée . . . . .	14
3.2 La fonction de répartition empirique bidimensionnelle . . . . .	18
3.3 La copule empirique . . . . .	19
<b>4 Covariance asymptotique</b>	<b>23</b>
4.1 Cas bidimensionnel . . . . .	24
4.2 Cas multidimensionnel . . . . .	26
4.3 Étude numérique . . . . .	30
4.3.1 Copule de Clayton . . . . .	31
4.3.2 Copule de Frank . . . . .	33
4.3.3 Copule de Gumbel–Hougaard . . . . .	34
<b>5 Conclusion</b>	<b>38</b>

BIBLIOGRAPHIE . . . . .	39
A Code R	40

# Table des figures

4.1	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Clayton de paramètre $\theta = 1$ . . . . .	32
4.2	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Clayton de paramètre $\theta = 10$ . . . . .	32
4.3	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Clayton de paramètre $\theta = 40$ . . . . .	33
4.4	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Frank de paramètre $\theta = 1$ . . . . .	34
4.5	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Frank de paramètre $\theta = 10$ . . . . .	35
4.6	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Frank de paramètre $\theta = 40$ . . . . .	35
4.7	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Gumbel–Hougaard de paramètre $\theta = 1$ . . . . .	36
4.8	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Gumbel–Hougaard de paramètre $\theta = 10$ . . . . .	36
4.9	Graphe des paires $(i, \mathcal{D}_{\theta(i)})$ pour $i \in \{1, \dots, 1\,000\,000\}$ pour la copule de Gumbel–Hougaard de paramètre $\theta = 40$ . . . . .	37

# Chapitre 1

## Introduction

La théorie des copules, issue des travaux de Sklar (1959), permet une modélisation flexible de la dépendance entre deux ou plusieurs variables aléatoires. Depuis quelques années, la croissance de l'intérêt pour cette théorie est phénoménale. Dans son article intitulé *Copulas : Tales and Facts*, le professeur Mikosch (2007) indiquait déjà qu'en septembre 2005, une recherche sur Google portant sur le terme *copula* produisait 650 000 résultats. En août 2015, cette même requête en engendre près de 800 000. À en juger par le nombre de publications scientifiques et autres travaux disponibles sur Internet, l'engouement envers la théorie des copules est plus grand que jamais.

Soit  $(X, Y)$  un couple de variables aléatoires continues dont la fonction de répartition bidimensionnelle, définie en tous  $x, y \in \mathbb{R}$  par  $H(x, y) = \Pr(X \leq x, Y \leq y)$ , a pour marges  $F(x) = \Pr(X \leq x)$  et  $G(y) = \Pr(Y \leq y)$ . Posons  $U = F(X)$  et  $V = G(Y)$ . La fonction de répartition du couple  $(U, V)$ , dont les marges sont uniformes sur l'intervalle  $(0, 1)$ , est appelée une copule. Cette fonction  $C$  est la seule qui permette d'exprimer la loi  $H$  sous la forme  $H(x, y) = C\{F(x), G(y)\}$  en tous  $x, y \in \mathbb{R}$ .

Puisque la copule  $C$  régit la dépendance entre les variables  $X$  et  $Y$ , son estimation est d'intérêt. Considérons pour ce faire un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  de loi  $H$ . Lorsque les marges  $F$  et  $G$  sont connues, on peut calculer  $U_i = F(X_i)$  et  $V_i = G(Y_i)$  pour tout  $i \in \{1, \dots, n\}$ , ce qui permet alors d'estimer  $C$  par la fonction de répartition empirique de l'échantillon  $(U_1, V_1), \dots, (U_n, V_n)$ . Il découle du théorème de Donsker (1952) que le processus empirique  $\mathbb{C}_n = \sqrt{n}(C_n - C)$  converge en loi vers un processus gaussien  $\mathbb{C}$  dont la fonction de covariance est donnée, en tous  $u, v, s, t \in [0, 1]$ , par

$$\text{cov}\{\mathbb{C}(u, v), \mathbb{C}(s, t)\} = C(u \wedge s, v \wedge t) - C(u, v)C(s, t),$$

où  $a \wedge b = \min(a, b)$ .

En pratique, toutefois, les lois marginales  $F$  et  $G$  sont le plus souvent inconnues. On peut alors les estimer au moyen de leurs versions empiriques  $F_n$  et  $G_n$ , ce qui conduit à estimer la copule  $C$  par la fonction de répartition  $\hat{C}_n$  du pseudo-échantillon  $(\hat{U}_1, \hat{V}_1), \dots, (\hat{U}_n, \hat{V}_n)$ , où pour tout  $i \in \{1, \dots, n\}$  on pose  $\hat{U}_i = F_n(X_i)$  et  $\hat{V}_i = G_n(Y_i)$ . Le processus empirique correspondant est défini par  $\hat{\mathbb{C}}_n = \sqrt{n}(\hat{C}_n - C)$ .

Le comportement limite du processus de copule empirique  $\hat{\mathbb{C}}_n$  a été étudié sous diverses conditions. Voir par exemple Rüschendorf (1976), Gänßler & Stute (1987), Fermanian et coll. (2004), Genest & Remillard (2004) et Segers (2012). Sa limite, dénotée  $\hat{\mathbb{C}}$ , est gaussienne mais différente de celle de  $\mathbb{C}$ . Genest & Segers (2010) ont comparé les fonctions de covariance associées à ces deux limites. Ils en ont tiré deux résultats principaux. Le premier stipule que sous certaines conditions de dépendance positive, la covariance du processus  $\hat{\mathbb{C}}$  est plus petite que celle du processus  $\mathbb{C}$ . Le second est que cette inégalité reste vraie en dimension  $d \geq 3$  pour la copule d'indépendance, définie en tous  $u_1, \dots, u_d \in (0, 1)$  par  $\Pi(u_1, \dots, u_d) = u_1 \times \dots \times u_d$ .

L'objectif de ce mémoire est d'explorer dans quelle mesure l'inégalité obtenue par Genest & Segers (2010) s'étend au-delà de l'indépendance en dimension  $d \geq 3$ . Le principal résultat théorique rapporté ici, dû aux professeurs Genest et Mesfioui, montre que tel est bien le cas au voisinage de l'indépendance et pour certains types de dépendance positive. Ces résultats sont décrits en détail et complétés par une étude numérique qui laisse présager que l'inégalité reste valide pour les familles paramétriques de copules archimédiennes de Clayton, Frank et Gumbel–Hougaard en dimension  $d = 4$ .

Le mémoire est structuré comme suit. Des notions générales sur les copules sont d'abord exposées au chapitre 2, où quelques exemples de copules paramétriques sont également donnés, ainsi qu'un algorithme générique de simulation d'une copule bivariée. Au chapitre 3, nous rappelons comment estimer les fonctions de répartition univariées et bivariées par leurs versions empiriques ; la notion de copule empirique y est également définie. Au chapitre 4, nous décrivons les principaux résultats de Genest & Segers (2010), ainsi que certaines de leurs généralisations encore inédites. Nous présentons ensuite les résultats des essais numériques que nous avons réalisés en complément de ces travaux. Après une brève conclusion, une annexe regroupant notre code R vient clore le mémoire.



# Chapitre 2

## Notions théoriques sur les copules

Le concept de copule joue un rôle primordial dans la modélisation de la dépendance entre les variables aléatoires. Dans ce chapitre, nous souhaitons rappeler quelques généralités sur cette théorie. Les définitions de base et quelques exemples classiques de copules sont donnés à la section 2.1. Nous présentons ensuite à la section 2.2 certains modèles de copules importants, ainsi que quelques-unes de leurs propriétés. La section 2.3 est consacrée à la classe des copules archimédiennes. À la section 2.4, nous présentons un algorithme générique de simulation d'une copule bivariée. À la section ??, nous illustrons cet algorithme pour différents modèles de copules. Les résultats présentés ici sont tirés de Nelsen (2006) et Genest & Nešlehová (2012).

### 2.1 Généralités sur les copules

Soit  $(X, Y)$  un couple de variables aléatoires dont la fonction de répartition conjointe est définie en tout  $(x, y) \in \mathbb{R}^2$  par

$$H(x, y) = \Pr(X \leq x, Y \leq y).$$

Les lois marginales des variables  $X$  et  $Y$  sont obtenues et notées comme suit :

$$\begin{aligned} F(x) &= \Pr(X \leq x) = \lim_{y \rightarrow \infty} H(x, y), \\ G(y) &= \Pr(Y \leq y) = \lim_{x \rightarrow \infty} H(x, y). \end{aligned}$$

### 2.1.1 Théorème de Sklar et propriétés

La théorie des copules permet de modéliser la loi conjointe  $H$  du couple  $(X, Y)$  lorsque celle-ci est inconnue. Très populaire depuis une trentaine d'années, cette approche permet de construire des modèles pour  $H$  sans pour autant contraindre le choix des marges  $F$  et  $G$ . Le point de départ de la théorie est le résultat suivant (Sklar, 1959).

**Théorème 2.1.1.** *Si  $H$  est une fonction de répartition bivariable de marges  $F$  et  $G$ , alors il existe une copule  $C : [0, 1]^2 \rightarrow [0, 1]$  telle que pour tout  $(x, y) \in \mathbb{R}^2$ ,*

$$H(x, y) = C\{F(x), G(y)\}. \quad (2.1)$$

*De plus, cette copule est unique lorsque les marges  $F$  et  $G$  sont continues.*

Ce résultat met en relief le fait que la loi  $H$  d'un couple  $(X, Y)$  est composée de trois éléments : les lois marginales  $F$  et  $G$ , ainsi qu'une fonction  $C$  appelée copule qui les lie. Tandis que les fonctions  $F$  et  $G$  décrivent le comportement marginal de  $X$  et de  $Y$ , leur dépendance est encodée dans la copule  $C$ .

Les propriétés caractérisant la fonction  $C$  apparaissant dans (2.1) sont les suivantes.

**Proposition 2.1.2.** *Pour qu'une fonction  $C : [0, 1]^2 \rightarrow [0, 1]$  soit une copule, il faut et il suffit qu'elle réponde aux critères suivants :*

- a) *pour tout  $u \in [0, 1]$ , on a  $C(u, 0) = C(0, u) = 0$  et  $C(u, 1) = C(1, u) = u$  ;*
- b) *pour tous  $u_1, u_2, v_1, v_2 \in [0, 1]$  tels que  $u_1 \leq u_2$  et  $v_1 \leq v_2$ , on a*

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Le résultat suivant précise que dans le cas où les marges  $F$  et  $G$  sont continues, la copule  $C$  est en fait la loi conjointe du couple  $(U, V)$ , où  $U = F(X)$  et  $V = G(Y)$ .

**Proposition 2.1.3.** *Soit  $H$  la fonction de répartition d'un couple  $(X, Y)$ . Si les marges  $F$  de  $X$  et  $G$  de  $Y$  sont continues, alors la copule  $C$  définie en (2.1) est la loi jointe du couple  $(U, V)$ , où  $U = F(X)$  et  $V = G(Y)$ .*

**Démonstration.** Pour tous  $u, v \in [0, 1]$ , on a

$$\begin{aligned} \Pr\{F(X) \leq u, G(Y) \leq v\} &= \Pr\{X \leq F^{-1}(u), Y \leq G^{-1}(v)\} \\ &= H\{F^{-1}(u), G^{-1}(v)\} \\ &= C\{F \circ F^{-1}(u), G \circ G^{-1}(v)\} \\ &= C(u, v), \end{aligned}$$

ce qu'il fallait démontrer. □

### 2.1.2 Utilité pour la modélisation

L'expression (2.1), appelée représentation de Sklar, est d'une grande utilité pour la modélisation du comportement stochastique d'une paire  $(X, Y)$  de variables aléatoires, car elle permet de décomposer la construction du modèle en trois étapes distinctes, à savoir (i) le choix de la marge  $F$  de  $X$  ; (ii) le choix de la marge  $G$  de  $Y$  ; et (iii) le choix de la copule  $C$  liant  $F$  et  $G$ .

Qui plus est, la structure de dépendance telle que représentée par la copule est invariante sous transformations croissantes des marges. Cette observation, capitale pour le développement de l'inférence, fait l'objet du résultat suivant.

**Proposition 2.1.4.** *Soient  $\kappa$  et  $\eta$  deux fonctions monotones croissantes. La copule du couple  $(X, Y)$  est la même que la copule du couple  $(\tilde{X}, \tilde{Y})$ , où  $\tilde{X} = \kappa(X)$  et  $\tilde{Y} = \eta(Y)$ .*

**Démonstration.** Pour tous  $x, y \in \mathbb{R}$ , on a

$$\begin{aligned} \tilde{H}(x, y) &= \Pr\{\kappa(X) \leq x, \eta(Y) \leq y\} \\ &= \Pr\{X \leq \kappa^{-1}(x), Y \leq \eta^{-1}(y)\} \\ &= H\{\kappa^{-1}(x), \eta^{-1}(y)\} \\ &= C\{F \circ \kappa^{-1}(x), G \circ \eta^{-1}(y)\}. \end{aligned}$$

Comme les lois marginales de  $\tilde{X}$  et  $\tilde{Y}$  sont respectivement  $F \circ \kappa^{-1}$  et  $G \circ \eta^{-1}$ , on déduit que la copule de  $\tilde{H}$  est  $C$ , ce qu'il fallait démontrer.  $\square$

Noter que si  $\kappa$  est monotone décroissante, la copule du couple  $(\tilde{X}, \tilde{Y})$  n'est toutefois pas la même que celle de  $(X, Y)$ . Un calcul simple montre qu'en fait, la copule de  $\tilde{H}$  est alors donnée en tous  $u, v \in (0, 1)$  par

$$\tilde{C}(u, v) = v - C(1 - u, v).$$

De même si  $\kappa$  est monotone croissante et  $\eta$  est monotone décroissante, alors la copule associée à la paire  $(\tilde{X}, \tilde{Y})$  est donnée en tous  $u, v \in (0, 1)$  par

$$\tilde{C}(u, v) = u - C(u, 1 - v).$$

Enfin si  $\kappa$  et  $\eta$  sont monotones décroissantes, alors  $\tilde{C}$  est la copule de survie associée à  $C$ , laquelle est donnée en tous  $u, v \in (0, 1)$  par

$$\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

### 2.1.3 Extraction d'une copule

Si la loi d'un couple  $(X, Y)$  est donnée en tous  $x, y \in \mathbb{R}$  par

$$H(x, y) = C\{F(x), G(y)\},$$

on peut alors déduire  $C$  de  $H$  en posant  $u = F(x)$  et  $v = G(y)$  dans l'équation précédente, ce qui permet d'écrire

$$C(u, v) = H\{F^{-1}(u), G^{-1}(v)\}.$$

L'exemple qui suit montre comment obtenir une copule à partir d'une fonction de répartition bivariée de marges continues.

**Exemple 2.1.5.** Soit  $H$  une fonction de répartition de paramètre  $\theta \geq 1$  définie comme suit en tous  $x, y \in \mathbb{R}$  :

$$H(x, y) = \exp\{-(e^{-\theta x} + e^{-\theta y})^{1/\theta}\}.$$

Les marges sont alors données en tous  $x, y \in \mathbb{R}$  par

$$F(x) = \lim_{y \rightarrow \infty} H(x, y) = \exp(-e^{-x}),$$

$$G(y) = \lim_{x \rightarrow \infty} H(x, y) = \exp(-e^{-y}).$$

Il s'ensuit que pour tous  $u, v \in (0, 1)$ , on a

$$F^{-1}(u) = -\ln\{-\ln(u)\}, \quad G^{-1}(v) = -\ln\{-\ln(v)\}.$$

La copule correspondante est donc définie en tous  $u, v \in (0, 1)$  par

$$C(u, v) = \exp[-\{\ln(1/v)^\theta + \ln(1/u)^\theta\}^{1/\theta}].$$

Toute copule est continue, mais pas nécessairement absolument continue. Lorsqu'une copule bivariée admet une densité, elle peut être calculée comme suit.

**Proposition 2.1.6.** La densité d'une copule, lorsque cette dernière existe, est donnée en tous  $u, v \in (0, 1)$  par

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v) = \frac{h\{F^{-1}(u), G^{-1}(v)\}}{f\{F^{-1}(u)\}g\{G^{-1}(v)\}},$$

où  $f, g, h$  dénotent respectivement les densités de  $F, G$  et  $H$ .

L'exemple qui suit illustre ce résultat.

**Exemple 2.1.7.** Soit  $H$  la fonction de répartition définie en tous  $x, y \in [0, 1]$  par

$$H(x, y) = \sqrt{\frac{xy}{x + y - xy}}.$$

La copule correspondante est définie en tous  $u, v \in (0, 1)$  par

$$C(u, v) = (u^{-2} + v^{-2} - 1)^{-1/2}.$$

La densité associée est donc donnée en tous  $u, v \in (0, 1)$  par

$$c(u, v) = 3(uv)^{-3}(u^{-2} + v^{-2} - 1)^{-5/2}.$$

## 2.1.4 Copule d'indépendance et bornes de Fréchet–Hoeffding

Une des copules les plus communes est celle associée à l'indépendance entre deux variables aléatoires. Comme l'indépendance est caractérisée par la factorisation de la loi jointe associée  $H$  en le produit de ses marges  $F$  et  $G$ , à savoir  $H = FG$ , une application directe du théorème de Sklar entraîne que la copule associée est définie en tous  $u, v \in [0, 1]$  par

$$\Pi(u, v) = uv.$$

Cette copule, qui caractérise l'indépendance bivariée, est aussi appelée copule produit.

La proposition suivante précise la plage de valeurs que peut prendre une copule arbitraire en un point arbitraire  $(u, v) \in [0, 1]^2$ .

**Proposition 2.1.8.** Soit  $C$  une copule bivariée. Pour tous  $u, v \in [0, 1]$ , on a

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq M(u, v) = \min(u, v).$$

Donc si  $H$  est une loi bivariée de marges  $F$  et  $G$ , alors pour tous  $x, y \in \mathbb{R}$  on a

$$\max\{F(x) + G(y) - 1, 0\} \leq H(x, y) \leq \min\{F(x), G(y)\}.$$

**Démonstration.** Soient deux événements quelconques  $A$  et  $B$ . On a

$$\Pr(A \cap B) \leq \Pr(A) \quad \text{et} \quad \Pr(A \cap B) \leq \Pr(B).$$

Par conséquent,

$$\Pr(A \cap B) \leq \min\{\Pr(A), \Pr(B)\}.$$

Étant donné  $x, y \in \mathbb{R}$ , appliquons cette inégalité aux événements  $A = \{X \leq x\}$  et  $B = \{Y \leq y\}$ . On trouve alors que

$$\Pr(X \leq x, Y \leq y) \leq \min\{\Pr(X \leq x), \Pr(Y \leq y)\},$$

ce qui revient à dire que  $H(x, y) \leq \min\{F(x), G(y)\}$ , tel qu'annoncé.

Par ailleurs, en utilisant les lois de De Morgan, on trouve que

$$\begin{aligned} \Pr(A \cap B) &= 1 - \Pr(A^c \cup B^c) \\ &= 1 - \{\Pr(A^c) + \Pr(B^c) - \Pr(A^c \cap B^c)\} \\ &= \Pr(A) + \Pr(B) - 1 + \Pr(A^c \cap B^c) \\ &\geq \Pr(A) + \Pr(B) - 1. \end{aligned}$$

Sachant que  $\Pr(A \cap B) \geq 0$ , on a donc

$$\Pr(A \cap B) \geq \max\{\Pr(A) + \Pr(B) - 1, 0\}.$$

Dans le cas particulier où  $A = \{X \leq x\}$  et  $B = \{Y \leq y\}$ , on déduit que

$$H(x, y) \geq \max\{F(x) + G(y) - 1, 0\},$$

ce qui est le résultat annoncé.  $\square$

Les fonctions  $W$  et  $M$  sont appelées respectivement borne inférieure et supérieure de Fréchet–Hoeffding. Puisque ce sont des copules, la proposition 2.1.8 ne peut être améliorée. De fait, la copule  $M$  correspond à la dépendance positive parfaite, c'est-à-dire au cas d'un vecteur  $(X, Y)$  où  $Y = \kappa(X)$  pour une fonction monotone croissante  $\kappa$ . En effet, comme on a alors  $G(y) = F \circ \kappa^{-1}(y)$  pour tout  $y \in \mathbb{R}$ , il s'ensuit que

$$V = G(Y) = F \circ \kappa^{-1}\{\kappa(X)\} = F(X) = U$$

de sorte que pour tous  $u, v \in [0, 1]$ , on trouve

$$\Pr(U \leq u, V \leq v) = \Pr(U \leq u, U \leq v) = \Pr\{U \leq \min(u, v)\} = \min(u, v).$$

Quant à  $W$ , c'est la copule de la dépendance négative parfaite, c'est-à-dire qu'elle correspond au cas d'un vecteur  $(X, Y)$  où  $Y = \kappa(X)$  pour une fonction monotone décroissante  $\kappa$ . En effet, comme  $G(y) = 1 - F \circ \kappa^{-1}(y)$  pour tout  $y \in \mathbb{R}$ , il vient

$$V = G(Y) = 1 - F \circ \kappa^{-1}\{\kappa(X)\} = 1 - F(X) = 1 - U$$

et donc pour tous  $u, v \in [0, 1]$ , trouve

$$\Pr(U \leq u, V \leq v) = \Pr(U \leq u, 1 - U \leq v) = \Pr(1 - v \leq U \leq u) = \max(u + v - 1, 0).$$

### 2.1.5 Cas multidimensionnel

Nous allons voir à présent que plusieurs des relations et propriétés présentées jusqu'ici peuvent s'étendre facilement au cas multidimensionnel. Commençons par énoncer une version multivariée du théorème de Sklar.

**Théorème 2.1.9.** *Soit un vecteur  $X = (X_1, \dots, X_d)$  de fonction de répartition  $H$ . Notons par  $F_1, \dots, F_d$  les fonctions de répartition marginales. Alors il existe une copule  $C : [0, 1]^d \rightarrow [0, 1]$  telle que pour tout  $(x_1, \dots, x_d) \in \mathbb{R}^d$ , on a la représentation*

$$H(x) = C\{F_1(x_1), \dots, F_d(x_d)\}.$$

*De plus, cette copule est unique lorsque les marges  $F_1, \dots, F_d$  sont continues.*

La fonction  $C : [0, 1]^d \rightarrow [0, 1]$  est appelée une copule si et seulement si elle possède les propriétés suivantes :

- a)  $C(u_1, \dots, u_d) = 0$  si au moins une des composantes  $u_1, \dots, u_d$  est 0 ;
- b)  $C(u_1, \dots, u_d) = u_i$  si  $u_j = 1$  pour tout  $j \in \{1, \dots, d\} \setminus \{i\}$  ;
- c) la mesure induite par  $C$  de tout rectangle de  $[0, 1]^d$  doit être non négative.

Les composantes de  $(X_1, \dots, X_d)$  sont indépendantes si et seulement si pour tous  $x_1, \dots, x_d \in \mathbb{R}$ , on a

$$\Pr(X_1 \leq x_1, \dots, X_d \leq x_d) = \prod_{i=1}^d \Pr(X_i \leq x_i).$$

La copule d'indépendance multivariée est donc donnée en tous  $u_1, \dots, u_d \in [0, 1]$  par

$$\Pi(u_1, \dots, u_d) = \prod_{i=1}^d u_i.$$

## 2.2 Quelques modèles de copule

Nous allons maintenant présenter quelques modèles de copules bien connus. En particulier, nous décrirons dans cette section la copule gaussienne, la copule de Farlie–Gumbel–Morgenstern et la copule de valeurs extrêmes de Gumbel–Hougaard.

Au moyen des copules, il est facile de construire un modèle dont la dépendance est de type normal bien que les marges soient arbitraires. Il suffit pour ce faire que la copule soit gaussienne. Les copules membres de cette classe sont définies ci-dessous.

**Exemple 2.2.1.** La copule gaussienne bidimensionnelle de paramètre  $\rho \in (-1, 1)$  est définie en tous  $u, v \in (0, 1)$  par

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} h_\rho(s, t) dt ds,$$

où

$$h_\rho(s, t) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)} \right\}$$

est la densité d'une loi normale à deux variables de moyennes nulles, de variances unitaires et de corrélation  $\rho$ . Ici,  $\Phi$  est la fonction de répartition d'une variable  $\mathcal{N}(0, 1)$ .

**Exemple 2.2.2.** Le modèle bivarié de Farlie–Gumbel–Morgenstern de paramètre  $\theta \in (-1, 1)$  est défini en tous  $x, y \in \mathbb{R}$  par

$$H_\theta(x, y) = F(x)G(y) + \theta F(x)G(y) \{1 - F(x)\} \{1 - G(y)\}.$$

Ce modèle a souvent été utilisé pour bâtir des contre-hypothèses dans l'étude de l'efficacité de tests d'indépendance. En effet, quand  $\theta = 0$  on retrouve l'indépendance. Quand  $\theta \neq 0$ , le second terme de la somme crée un écart à l'indépendance et permet donc l'étude de la puissance de tests sous ces contre-hypothèses. La copule extraite de  $H_\theta$  est la copule de Farlie–Gumbel–Morgenstern, définie en tous  $u, v \in [0, 1]$  par

$$C_w(uv) = uv - \theta uv(1 - u)(1 - v).$$

Il existe bien d'autres classes de copules, dont les copules de valeurs extrêmes.

**Exemple 2.2.3.** Les copules de valeurs extrêmes permettent de modéliser des phénomènes extrêmes tels les ouragans ou la crue des eaux. Une copule extrême, notée  $C_A$ , est la fonction de dépendance extraite d'une loi bivariée extrême. Elle s'exprime en tous  $u, v \in (0, 1)$  sous la forme

$$C_A(u, v) = \exp \left\{ \ln(uv) A \left( \frac{\ln v}{\ln u + \ln v} \right) \right\}.$$

où  $A : [0, 1] \rightarrow [1/2, 1]$  est une fonction convexe, appelée fonction de dépendance de Pickands, qui doit vérifier la condition  $\max(t, 1 - t) \leq A(t) \leq 1$  en tout  $t \in [0, 1]$ .

Les copules de valeurs extrêmes ne permettent de modéliser qu'une dépendance positive car pour tous  $u, v \in [0, 1]$ , elles vérifient la condition

$$uv \leq C_A(u, v) \leq \min(u, v).$$

Lorsque  $A(t) = 1$  en tout  $t \in [0, 1]$ , on retrouve la copule d'indépendance. Ainsi, la copule  $\Pi(u, v) = uv$  est un modèle extrême. Les copules dites de Gumbel–Hougaard appartiennent aussi à cette classe. Ces copules dépendent d'un paramètre  $\theta \geq 1$  et sont engendrées par la fonction de Pickands donnée en tout  $t \in (0, 1)$  par la formule

$$A_\theta(t) = \{t^\theta + (1 - t)^\theta\}^{1/\theta}.$$



## 2.3 La famille des copules archimédiennes

Cette section est consacrée à une classe générale de copules appelées copules archimédiennes, introduite par Genest & MacKay (1986).

**Définition 2.3.1.** *On dit qu'une copule bidimensionnelle est archimédienne si elle s'exprime en tous  $u, v \in (0, 1)$  sous la forme*

$$C_\phi(u, v) = \phi^{-1} \{ \phi(u) + \phi(v) \}. \quad (2.2)$$

*en terme d'une fonction  $\phi : [0, 1] \rightarrow [0, \infty)$  décroissante et convexe telle que  $\phi(1) = 0$ . Cette fonction est appelée un générateur de  $C_\phi$ .*

**Proposition 2.3.1.** *Soit  $C_\phi$  une copule archimédienne quelconque. Alors*

- (i) *pour tous  $u, v \in (0, 1)$ , on a  $C_\phi(u, v) = C_\phi(v, u)$  ;*
- (ii) *si  $\phi^*(t) = k\phi(t)$  pour un certain  $k > 0$  et tout  $t \in (0, 1)$ , alors  $C_{\phi^*} = C_\phi$  ;*
- (iii)  *$C_\phi$  est la copule produit si et seulement si  $\phi(t) \propto -\ln t$  en tout  $t \in (0, 1)$ .*

Les copules archimédiennes peuvent être généralisées au cas multidimensionnel. Cette extension est définie en tous  $u_1, \dots, u_d \in (0, 1)$  par

$$C(u_1, \dots, u_d) = \phi^{-1} \{ \phi(u_1) + \dots + \phi(u_d) \},$$

où  $\phi(1) = 0$  et  $\phi^{-1}$  est une fonction  $d$ -monotone, au sens précisé dans l'article de McNeil & Nešlehová (2009). Si on exige qu'une fonction  $\phi$  engendre une copule en toute dimension  $d \geq 2$ , il faut alors que  $\phi$  soit complètement monotone, c'est-à-dire que pour tout  $i \in \mathbb{N}$ , on ait

$$(-1)^i \frac{d^i}{dt^i} \phi^{-1}(t) > 0.$$

Plusieurs familles de copules connues appartiennent à la classe des copules archimédiennes, notamment les copules de Clayton, de Frank et de Gumbel-Hougaard.

**Exemple 2.3.2.** *Le générateur de la copule de Clayton de paramètre  $\theta > 0$  est donné en tout  $t \in (0, 1)$  par*

$$\phi_\theta(t) = \frac{t^{-\theta} - 1}{\theta}.$$

*En substituant cette formule dans l'expression (2.2), on trouve*

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$$

*en tous  $u, v \in (0, 1)$ . De plus,*

$$\lim_{\theta \rightarrow 0} C_\theta(u, v) = uv,$$

*de sorte que la copule d'indépendance correspond au cas limite où  $\theta \rightarrow 0$ .*

**Exemple 2.3.3.** Le générateur de la copule de Frank est donné en tout  $t \in (0, 1)$  par

$$\phi_\theta(t) = \ln \left( \frac{1 - e^{-\theta}}{1 - e^{-\theta t}} \right),$$

où  $\theta \neq 0$ . La copule de Frank elle-même est donnée en tous  $u, v \in (0, 1)$  par

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left\{ 1 - \frac{(1 - e^{-\theta u})(1 - e^{-\theta v})}{1 - e^{-\theta}} \right\}.$$

**Exemple 2.3.4.** Le générateur de la copule de Gumbel-Hougaard est donné en tout  $t \in (0, 1)$  par

$$\phi_\theta(t) = \{-\ln(t)\}^\theta,$$

où  $\theta \geq 1$ . La copule de Gumbel-Hougaard est donnée en tous  $u, v \in (0, 1)$  par

$$C_\theta(u, v) = \exp \left[ - \left\{ (-\ln u)^\theta + (-\ln v)^\theta \right\}^{1/\theta} \right].$$

Noter que le cas  $\theta = 1$  correspond à l'indépendance.

## 2.4 Simulation de copules bivariées

**Théorème 2.4.1.** Soit  $C$  une copule et soit  $(U, V)$  un vecteur aléatoire de loi  $C$ . On définit pour tout  $(u, v) \in (0, 1)^2$

$$L_v(u) = \Pr(U \leq u | V = v) = \frac{\partial}{\partial v} C(u, v).$$

Soient maintenant  $T_1$  et  $T_2$  deux variables aléatoires indépendantes telle que  $T_1 \sim \mathcal{U}_{(0,1)}$  et  $T_2 \sim \mathcal{U}_{(0,1)}$ . Si  $U = L_{T_2}^{-1}(T_1)$  et  $V = T_2$ , alors la paire  $(U, V)$  est de loi  $C$ .

**Démonstration.** Pour tous  $u, v \in (0, 1)$ , on a

$$\begin{aligned} \Pr(U \leq u, V \leq v) &= \Pr\{L_{T_2}^{-1}(T_1) \leq u, T_2 \leq v\} = \int_0^v \Pr\{L_{T_2}^{-1}(T_1) \leq u | T_2 = s\} ds \\ &= \int_0^v \Pr\{L_s^{-1}(T_1) \leq u\} ds = \int_0^v \Pr\{T_1 \leq L_s(u)\} ds = \int_0^v L_s(u) ds \\ &= \int_0^v \frac{\partial}{\partial s} C(u, s) ds = C(u, v), \end{aligned}$$

ce qui montre que la loi du couple  $(U, V)$  est bel et bien  $C$ . □

On peut donc établir un algorithme de simulation de copule comme suit :

- a) On génère des variables aléatoires et indépendantes  $V, T$  qui sont  $\mathcal{U}_{(0,1)}$ .
- b) On pose  $U = L_V^{-1}(T)$ .
- c) Le couple désiré est  $(U, V)$ .

Appliquons cet algorithme dans le cas d'une copule archimédienne  $C$  de générateur  $\phi$ . On a vu que  $C$  s'écrit en tous  $u, v \in (0, 1)$  sous la forme  $C_\phi(u, v) = \phi^{-1} \{ \phi(u) + \phi(v) \}$ . Par définition même de  $L_v(u)$ , on a

$$L_v(u) = \frac{\partial}{\partial v} C(u, v) = \frac{\phi'(v)}{\phi'[\phi^{-1} \{ \phi(u) + \phi(v) \}]}.$$

Ainsi, l'inverse de  $L_v(u)$  est donné en tout  $t \in (0, 1)$  par

$$L_v^{-1}(t) = \phi^{-1} \left[ \phi \left[ (\phi')^{-1} \left\{ \frac{\phi'(v)}{t} \right\} \right] - \phi(v) \right].$$

Pour tirer une observation  $(U, V)$  d'une copule archimédienne  $C_\phi$ , on peut donc procéder comme suit :

- a) On génère des variables aléatoires et indépendantes  $V, T$  qui sont uniformes sur l'intervalle  $(0, 1)$ .
- b) On pose

$$U = L_V^{-1}(T) = \phi^{-1} \left[ \phi \left[ (\phi')^{-1} \left\{ \frac{\phi'(V)}{T} \right\} \right] - \phi(V) \right].$$

Le couple  $(U, V)$  est alors de loi  $C_\phi$ .

**Exemple 2.4.2.** *L'algorithme de simulation pour une copule de Clayton de paramètre  $\theta > 0$  est le suivant :*

- a) On génère des variables aléatoires et indépendantes  $V, T$  qui sont uniformes sur l'intervalle  $(0, 1)$ .
- b) On pose

$$U = \left( V^{-\theta} T^{-\frac{\theta}{\theta+1}} - V^{-\theta} + 1 \right)^{-1/\theta}.$$

**Exemple 2.4.3.** *L'algorithme de simulation pour une copule de Frank de paramètre  $\theta \in \mathbb{R}$  est le suivant :*

- a) On génère des variables aléatoires et indépendantes  $V, T$  qui sont uniformes sur l'intervalle  $(0, 1)$ .
- b) On pose

$$U = \frac{1}{\theta} \ln \left\{ \frac{e^{-\theta V} + T(1 - e^{-\theta V})}{e^{-\theta V} + T(e^{-\theta} - e^{-\theta V})} \right\}.$$

# Chapitre 3

## Processus empiriques

Nous avons défini au chapitre 2 la notion de copule ainsi que quelques-unes de ses propriétés fondamentales. Nous avons souligné entre autres que les copules jouent un rôle très important dans la modélisation de vecteurs aléatoires. Dans ce troisième chapitre, nous nous penchons maintenant sur la question de l'estimation d'une copule à partir d'un échantillon aléatoire de loi continue.

Nous commençons par présenter à la section 3.1 la notion de fonction de répartition empirique univariée. Nous introduisons ensuite à la section 3.2 la fonction de répartition empirique bidimensionnelle et nous expliquons en quoi celle-ci permet d'estimer une copule lorsque les lois marginales des vecteurs d'observations sont connues. Puis, à la section 3.3, nous montrons comment estimer une copule dans le cas plus commun où les marges sont inconnues. Pour simplifier la notation, les concepts et calculs sont limités au cas bivarié, mais la généralisation au cas multidimensionnel est immédiate.

### 3.1 La fonction de répartition empirique univariée

Soit  $X_1, \dots, X_n$  un échantillon aléatoire d'une loi  $F$ . La fonction de répartition empirique associée à cet échantillon est alors définie, en tout  $x \in \mathbb{R}$ , par

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x),$$

où, pour chaque  $i \in \{1, \dots, n\}$ , la fonction indicatrice est définie par

$$\mathbf{1}(X_i \leq x) = \begin{cases} 1 & \text{si } X_i \leq x; \\ 0 & \text{autrement.} \end{cases}$$

Pour un  $x \in \mathbb{R}$  fixé,  $F_n(x)$  correspond à la proportion des observations qui sont inférieures ou égales à  $x$ . La variable aléatoire  $nF_n(x)$  est donc une variable binomiale de paramètres  $n$  et  $F(x)$ . En particulier, on a

$$\mathbb{E}\{F_n(x)\} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\mathbf{1}_{(-\infty, x]}(X_i)\} = \frac{1}{n} \sum_{i=1}^n \Pr(X_i \leq x) = F(x)$$

et

$$\begin{aligned} \text{var}\{F_n(x)\} &= \frac{1}{n^2} \sum_{i=1}^n \text{var}\{\mathbf{1}_{(-\infty, x]}(X_i)\} \\ &= \frac{1}{n^2} \sum_{i=1}^n \Pr(X_i \leq x) \{1 - \Pr(X_i \leq x)\} \\ &= \frac{F(x) \{1 - F(x)\}}{n}. \end{aligned}$$

Puisque  $F_n(x)$  est une estimation sans biais de  $F(x)$  et que  $\text{var}\{F_n(x)\} \rightarrow 0$  quand  $n \rightarrow \infty$ , il est clair que  $F_n(x) \rightarrow F(x)$  en probabilité. Autrement dit,  $F_n(x)$  est une estimation convergente de  $F(x)$ , et ce pour tout  $x \in \mathbb{R}$ . De fait, la convergence est uniforme, comme l'ont montré Glivenko et Cantelli. Ce résultat classique (Billingsley 2012, p. 268) est rappelé ci-dessous.

**Théorème 3.1.1.** *Quand  $n \rightarrow \infty$ ,  $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0$  presque sûrement.*

On peut donc voir la *fonction*  $F_n$  comme un estimateur uniformément convergent de la fonction  $F$ . Évidemment, les valeurs prises par  $F_n$  en différents points  $x, x' \in \mathbb{R}$  sont dépendantes, puisque  $F_n$  est croissante. Il n'est pas étonnant, dès lors, que les variables aléatoires  $F_n(x)$  et  $F_n(x')$  soient dépendantes. En effet, on a

$$\begin{aligned} \text{cov}\{F_n(x), F_n(x')\} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov}\{\mathbf{1}_{(-\infty, x]}(X_i), \mathbf{1}_{(-\infty, x']} (X_j)\} \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{cov}\{\mathbf{1}_{(-\infty, x]}(X_i), \mathbf{1}_{(-\infty, x']} (X_i)\}, \end{aligned}$$

puisque les variables  $X_i$  et  $X_j$  sont indépendantes lorsque  $i \neq j$ . Il en va donc de même des variables de Bernoulli  $\mathbf{1}_{(-\infty, x]}(X_i)$  et  $\mathbf{1}_{(-\infty, x']} (X_j)$ . De plus,

$$\mathbb{E}\{\mathbf{1}_{(-\infty, x]}(X_i)\} = F(x), \quad \mathbb{E}\{\mathbf{1}_{(-\infty, x']} (X_i)\} = F(x')$$

et

$$\mathbb{E}\{\mathbf{1}_{(-\infty, x]}(X_i) \mathbf{1}_{(-\infty, x']} (X_i)\} = \Pr(X_i \leq x, X_i \leq x') = F(x \wedge x'),$$

où  $x \wedge x' = \min(x, x')$ . Ainsi,

$$\text{cov}\{\mathbf{1}_{(-\infty, x]}(X_i), \mathbf{1}_{(-\infty, x']} (X_i)\} = F(x \wedge x') - F(x)F(x')$$

et

$$\text{cov} \{F_n(x), F_n(x')\} = \frac{F(x \wedge x') - F(x)F(x')}{n}.$$

Dans le cas particulier où  $F$  est la fonction de répartition d'une variable uniforme sur l'intervalle  $(0, 1)$ , il s'ensuit que, pour tous  $x, x' \in (0, 1)$ , on a

$$\text{cov} \{F_n(x), F_n(x')\} = \frac{x \wedge x' - xx'}{n}.$$

Pour tout  $x \in \mathbb{R}$  fixé, considérons la variable aléatoire

$$\mathbb{F}_n(x) = \sqrt{n} \{F_n(x) - F(x)\}.$$

La variable  $\sqrt{n} \mathbb{F}_n(x)$  n'est rien d'autre qu'une variable aléatoire binomiale centrée et réduite. L'application du théorème de la limite centrale assure que quand  $n \rightarrow \infty$ ,

$$\mathbb{F}_n(x) \rightsquigarrow \mathcal{N}(0, \sigma_x^2),$$

où  $\rightsquigarrow$  dénote la convergence en loi et  $\sigma_x^2 = F(x) \{1 - F(x)\}$ . Ce résultat vaut pour tout  $x \in \mathbb{R}$ . De plus, si on fixe  $x, x' \in \mathbb{R}$ , la version bivariable du théorème de la limite centrale entraîne que quand  $n \rightarrow \infty$ ,

$$\begin{pmatrix} \mathbb{F}_n(x) \\ \mathbb{F}_n(x') \end{pmatrix} \rightsquigarrow \mathcal{N}_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{xx'} \right],$$

où

$$\Sigma_{xx'} = \begin{pmatrix} F(x)\{1 - F(x)\} & F(x \wedge x') - F(x)F(x') \\ F(x \wedge x') - F(x)F(x') & F(x')\{1 - F(x')\} \end{pmatrix}.$$

On pourrait énoncer un résultat semblable pour un triplet  $(F_n(x), F_n(x'), F_n(x''))$ , etc. De façon plus générale, il existe un théorème qui permet de préciser simultanément le comportement asymptotique de  $F_n$  comme fonction aléatoire en toute valeur  $x \in \mathbb{R}$ . Grosso modo, la limite de ce *processus empirique* est une sorte de “loi normale de dimension infinie” liée à la notion de “pont brownien”. Ce concept est défini ci-dessous.

**Définition 3.1.1.** *Un pont brownien est un processus gaussien  $\mathbb{B}$  sur  $[0, 1]$  tel que pour tous  $x, x' \in [0, 1]$ ,  $E\{\mathbb{B}(x)\} = 0$  et  $\text{cov} \{\mathbb{B}(x), \mathbb{B}(x')\} = x \wedge x' - xx'$ .*

Le résultat suivant est dû à Donsker (1952).

**Théorème 3.1.2.** *Soit  $D(\mathbb{R})$  l'espace des fonctions définies sur  $\mathbb{R}$  qui sont continues à droite et qui possèdent une limite à gauche. Quand  $n \rightarrow \infty$ , le processus empirique*

$$\mathbb{F}_n = \sqrt{n} (F_n - F) \rightsquigarrow \mathbb{F},$$

où  $\rightsquigarrow$  dénote la convergence en loi dans l'espace  $D(\mathbb{R})$  et la limite s'exprime sous la forme  $\mathbb{F} = \mathbb{B} \circ F$  en terme du pont brownien  $\mathbb{B}$  sur l'intervalle  $[0, 1]$ .

Au vu des propriétés de  $\mathbb{B}$  énoncées ci-dessus, on déduit que pour tous  $x, x' \in \mathbb{R}$ ,

$$\mathbb{E} \{ \mathbb{F}(x) \} = 0, \quad \text{cov} \{ \mathbb{F}(x), \mathbb{F}(x') \} = F(x \wedge x') - F(x)F(x').$$

Le théorème de Donsker est une généralisation très forte du théorème de la limite centrale, valable en fait en toute dimension finie.

À titre d'illustration, soit  $X_1, \dots, X_n$  des observations mutuellement indépendantes d'une variable  $X$  de loi  $F$  continue à support dans  $(0, \infty)$  et dénotons par  $F_n$  la fonction de répartition expérimentale correspondante. On a

$$\mu = \mathbb{E}(X) = \int_0^\infty \{1 - F(x)\} dx,$$

comme on peut facilement le montrer en utilisant le fait que pour tout  $x \in \mathbb{R}$ ,

$$x = \int_0^\infty \{1 - \mathbf{1}_{[x, \infty)}(t)\} dt.$$

Par analogie, on a aussi

$$\bar{X}_n = \int_0^\infty \{1 - F_n(x)\} dx.$$

Il s'ensuit que

$$\sqrt{n}(\bar{X}_n - \mu) = - \int_0^\infty \{ \mathbb{F}_n(x) \} dx.$$

Puisque  $\mathbb{F}_n \rightsquigarrow \mathbb{B} \circ F$  quand  $n \rightarrow \infty$ , on peut déduire du théorème des applications continues que

$$\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow - \int_0^\infty \mathbb{B} \{ F(x) \} dx.$$

Ainsi la limite de  $\sqrt{n}(\bar{X}_n - \mu)$  s'écrit comme l'intégrale de variables normales, dont la loi est elle-même normale. En effet, il découle du théorème central limite que

$$\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow \mathcal{N}(0, \sigma^2),$$

où  $\sigma^2 = \text{var}(X)$ . En combinant les deux résultats, on conclut que

$$\sigma^2 = \text{var} \left[ - \int_0^\infty \mathbb{B} \{ F(x) \} dx \right] \tag{3.1}$$

$$= \mathbb{E} \left[ \left[ - \int_0^\infty \mathbb{B} \{ F(x) \} dx \right]^2 \right] \tag{3.2}$$

$$= \mathbb{E} \left[ \int_0^\infty \int_0^\infty \mathbb{B} \{ F(x) \} \mathbb{B} \{ F(x') \} dx dx' \right] \tag{3.3}$$

$$= \int_0^\infty \int_0^\infty \mathbb{E} [\mathbb{B} \{ F(x) \} \mathbb{B} \{ F(x') \}] dx dx'. \tag{3.4}$$

En notant que

$$\mathbb{E} [\mathbb{B} \{ F(x) \} \mathbb{B} \{ F(x') \}] = F(x \wedge x') - F(x)F(x') = F(x) \wedge F(x') - F(x)F(x'),$$

on trouve

$$\sigma^2 = \int_0^\infty \int_0^\infty \{ F(x) \wedge F(x') - F(x)F(x') \} dx' dx,$$

identité qui n'est pas évidente a priori mais néanmoins vraie lorsque  $F(0) = 0$ .

## 3.2 La fonction de répartition empirique bidimensionnelle

Soit un échantillon aléatoire  $(X_1, Y_1), \dots, (X_n, Y_n)$  tiré d'une loi bivariable dont la fonction de répartition est  $H$ . On obtient une estimation de  $H$  par la fonction de répartition empirique bidimensionnelle définie en tous  $x, y \in \mathbb{R}$  par

$$H_n(x, y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x, Y_i \leq y).$$

Lorsque les variables  $x$  et  $y$  sont fixées,  $nH_n(x, y)$  est une variable binomiale de paramètres  $n$  et  $H(x, y)$ . On a donc

$$\mathbb{E}\{H_n(x, y)\} = H(x, y)$$

et

$$\text{var}\{H_n(x, y)\} = \frac{H(x, y)\{1 - H(x, y)\}}{n}.$$

Pour  $(x, y) \in \mathbb{R}^2$  fixé et  $\sigma_{x,y}^2 = H(x, y)\{1 - H(x, y)\}$ , le théorème de la limite centrale permet aussi d'affirmer que quand  $n \rightarrow \infty$ , on a

$$\mathbb{H}_n(x, y) = \sqrt{n}\{H_n(x, y) - H(x, y)\} \rightsquigarrow \mathcal{N}(0, \sigma_{x,y}^2).$$

Le résultat suivant est la version bivariable du Théorème 3.1.2, due à Donsker (1952).

**Théorème 3.2.1.** *Soit  $D(\mathbb{R}^2)$  l'espace des fonctions définies sur  $\mathbb{R}^2$  qui sont continues à droite et qui possèdent une limite à gauche en chacun de leur argument. Quand  $n \rightarrow \infty$ , le processus empirique*

$$\mathbb{H}_n = \sqrt{n}(H_n - H) \rightsquigarrow \mathbb{H},$$

*où  $\rightsquigarrow$  dénote la convergence en loi dans l'espace  $D(\mathbb{R}^2)$  et la limite est un drap brownien dont l'espérance est nulle et dont la fonction de covariance est donnée, en tous  $x, x', y, y' \in \mathbb{R}$  par*

$$\text{cov}\{\mathbb{H}(x, y), \mathbb{H}(x', y')\} = H(x \wedge x', y \wedge y') - H(x, y)H(x', y').$$

Le cas où les marges de  $H$  sont uniformes sur l'intervalle  $(0, 1)$  est d'un intérêt particulier, puisqu'il correspond à la situation où les paires  $(X_1, Y_1), \dots, (X_n, Y_n)$  sont issues d'une copule. Ce résultat peut être reformulé comme suit pour référence.



**Corollaire 3.2.2.** Soit  $C$  une copule et  $(U_1, V_1), \dots, (U_n, V_n)$  un échantillon aléatoire de loi  $C$ . Pour tous  $u, v \in [0, 1]$ , posons

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(U_i \leq u, V_i \leq v)$$

et  $\mathbb{C}_n(u, v) = \sqrt{n} \{C_n(u, v) - C(u, v)\}$ . Quand  $n \rightarrow \infty$ ,

$$\mathbb{C}_n(u, v) = \sqrt{n} (C_n - C) \rightsquigarrow \mathbb{C},$$

où  $\rightsquigarrow$  dénote la convergence en loi dans l'espace  $D([0, 1]^2)$  et la limite est un drap brownien dont l'espérance est nulle et dont la fonction de covariance est donnée, en tous  $u, u', v, v' \in [0, 1]^2$  par

$$\text{cov} \{ \mathbb{C}(u, v), \mathbb{C}(u', v') \} = C(u \wedge u', v \wedge v') - C(u, v)C(u', v').$$

Le Corollaire 3.2.2 nous assure que si l'on dispose d'un échantillon d'une copule  $C$ , la fonction  $C_n$  est alors un estimateur convergent et asymptotiquement gaussien de  $C$ . Dans la pratique, toutefois, on observe rarement un tel échantillon. Ce à quoi on a plutôt accès, c'est un échantillon de loi  $H$  dont la copule est  $C$ . Dans pareil cas, l'estimateur  $C_n$  est impraticable. Comment peut-on alors estimer  $C$ ? La réponse à cette question est donnée ci-dessous.

### 3.3 La copule empirique

Soit  $H$  une loi bivariée de marges continues  $F$  et  $G$  inconnues liées par une unique copule  $C$ . Étant donné un échantillon aléatoire  $(X_1, Y_1), \dots, (X_n, Y_n)$  de loi  $H$ , comment peut-on estimer  $C$ ?

Pour répondre à cette question, on part du fait — garanti par le théorème de Sklar — que la copule  $C$  associée à la loi  $H$  est la loi du couple  $(U, V) = (F(X), G(Y))$ . Si les marges  $F$  et  $G$  étaient connues, on pourrait estimer  $C$  par  $C_n$  après avoir posé

$$(U_1, V_1) = (F(X_1), G(Y_1)), \dots, (U_n, V_n) = (F(X_n), G(Y_n)).$$

Comme ceci s'avère impossible, on fonde plutôt l'estimation de  $C$  sur les pseudo-observations

$$(\hat{U}_1, \hat{V}_1) = (F_n(X_1), G_n(Y_1)), \dots, (\hat{U}_n, \hat{V}_n) = (F_n(X_n), G_n(Y_n)),$$

où  $F_n$  et  $G_n$  sont les fonctions de répartition expérimentales respectives de  $F$  et de  $G$ , définies en tous  $x, y \in \mathbb{R}$  par

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x) \quad \text{et} \quad G_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Y_i \leq y).$$

L'estimateur de  $C$  est alors défini en tout  $(u, v) \in [0, 1]^2$  par

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{U}_i \leq u, \hat{V}_i \leq v).$$

Cette fonction  $\hat{C}_n$  est appelée la “copule empirique”, bien qu’il ne s’agisse pas vraiment d’une copule ; en effet, les marges de la fonction  $\hat{C}_n$  ne sont pas uniformes sur  $(0, 1)$ .

Il est bon de noter que la copule empirique  $\hat{C}_n$  est une fonction des paires de rangs normés des observations. En effet, pour tout  $i \in \{1, \dots, n\}$ , soit  $R_i$  le rang de  $X_i$  parmi  $X_1, \dots, X_n$  et soit  $S_i$  le rang de  $Y_i$  parmi  $Y_1, \dots, Y_n$  ; alors

$$F_n(X_i) = R_i/n, \quad G_n(Y_i) = S_i/n.$$

Le résultat suivant caractérise le comportement asymptotique du processus de copule empirique défini en tout  $(u, v) \in [0, 1]^2$  par

$$\hat{\mathbb{C}}_n(u, v) = \sqrt{n} \{ \hat{C}_n(u, v) - C(u, v) \}.$$

Les conditions de régularité sous lesquelles le résultat est énoncé sont suffisantes mais pas forcément les plus faibles ; voir Segers (2012).

**Théorème 3.3.1.** *Soit  $C$  une copule dont les dérivées partielles  $C_{10}(u, v) = \partial C(u, v)/\partial u$  et  $C_{01}(u, v) = \partial C(u, v)/\partial v$  existent et sont continues pour tous  $u, v \in (0, 1)$ . Quand  $n \rightarrow \infty$ , le processus de copule empirique  $\hat{\mathbb{C}}_n = \sqrt{n}(\hat{C}_n - C)$  converge dans  $D([0, 1]^2)$  vers un processus gaussien  $\hat{\mathbb{C}}$  dont la représentation est donnée en tous  $u, v \in (0, 1)$  par*

$$\hat{\mathbb{C}}(u, v) = \mathbb{C}(u, v) - C_{10}(u, v)\mathbb{C}(u, 1) - C_{01}(u, v)\mathbb{C}(1, v).$$

Étant donné  $u, v, s, t \in (0, 1)$ , on a donc

$$\begin{aligned}
 \text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} &= \text{cov}\{\mathbb{C}(u, v) - C_{10}(u, v)\mathbb{C}(u, 1) - C_{01}(u, v)\mathbb{C}(1, v), \\
 &\quad \mathbb{C}(s, t) - C_{10}(s, t)\mathbb{C}(s, 1) - C_{01}(s, t)\mathbb{C}(1, t)\} \\
 &= \text{cov}\{\mathbb{C}(u, v), \mathbb{C}(s, t)\} - \text{cov}\{\mathbb{C}(u, v), C_{10}(s, t)\mathbb{C}(s, 1)\} \\
 &\quad - \text{cov}\{\mathbb{C}(u, v), C_{01}(s, t)\mathbb{C}(1, t)\} \\
 &\quad - \text{cov}\{C_{10}(u, v)\mathbb{C}(u, 1), \mathbb{C}(s, t)\} \\
 &\quad + \text{cov}\{C_{10}(u, v)\mathbb{C}(u, 1), C_{10}(s, t)\mathbb{C}(s, 1)\} \\
 &\quad + \text{cov}\{C_{10}(u, v)\mathbb{C}(u, 1), C_{01}(s, t)\mathbb{C}(1, t)\} \\
 &\quad - \text{cov}\{C_{01}(u, v)\mathbb{C}(1, v), \mathbb{C}(s, t)\} \\
 &\quad + \text{cov}\{C_{01}(u, v)\mathbb{C}(1, v), C_{10}(s, t)\mathbb{C}(s, 1)\} \\
 &\quad + \text{cov}\{C_{01}(u, v)\mathbb{C}(1, v), C_{01}(s, t)\mathbb{C}(1, t)\}.
 \end{aligned}$$

En faisant appel au Corollaire 3.2.2, on peut réduire cette expression à

$$\begin{aligned}
 \text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} &= C(u \wedge s, v \wedge t) - C(u, v)C(s, t) \\
 &\quad - C_{10}(s, t)\{C(u \wedge s, v \wedge 1) - C(u, v)C(s, 1)\} \\
 &\quad - C_{01}(s, t)\{C(u \wedge 1, v \wedge t) - C(u, v)C(1, t)\} \\
 &\quad - C_{10}(u, v)\{C(u \wedge s, 1 \wedge t) - C(u, 1)C(s, t)\} \\
 &\quad + C_{10}(u, v)C_{10}(s, t)\{C(u \wedge s, 1 \wedge 1) - C(u, 1)C(s, 1)\} \\
 &\quad + C_{10}(u, v)C_{01}(s, t)\{C(u \wedge 1, 1 \wedge t) - C(u, 1)C(1, t)\} \\
 &\quad - C_{01}(u, v)\{C(1 \wedge s, v \wedge t) - C(1, v)C(s, t)\} \\
 &\quad + C_{01}(u, v)C_{10}(s, t)\{C(1 \wedge s, v \wedge 1) - C(1, v)C(s, 1)\} \\
 &\quad + C_{01}(u, v)C_{01}(s, t)\{C(1 \wedge 1, v \wedge t) - C(1, v)C(1, t)\}.
 \end{aligned}$$

Après simplification, on trouve

$$\begin{aligned}
 \text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} &= C(u \wedge s, v \wedge t) - C(u, v)C(s, t) \\
 &\quad - C_{10}(s, t)\{C(u \wedge s, v) - sC(u, v)\} \\
 &\quad - C_{01}(s, t)\{C(u, v \wedge t) - tC(u, v)\} \\
 &\quad - C_{10}(u, v)\{C(u \wedge s, t) - uC(s, t)\} \\
 &\quad + C_{10}(u, v)C_{10}(s, t)(u \wedge s - us) \\
 &\quad + C_{10}(u, v)C_{01}(s, t)\{C(u, t) - ut\} \\
 &\quad - C_{01}(u, v)\{C(s, v \wedge t) - vC(s, t)\} \\
 &\quad + C_{01}(u, v)C_{10}(s, t)\{C(s, v) - vs\} \\
 &\quad + C_{01}(u, v)C_{01}(s, t)(v \wedge t - vt).
 \end{aligned}$$

**Exemple 3.3.2.** Nous nous intéressons à calculer la différence

$$\text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} - \text{cov}\{\mathbb{C}(u, v), \mathbb{C}(s, t)\}$$

pour la copule d'indépendance définie par  $C(u, v) = uv$  pour tout  $(u, v) \in [0, 1]^2$ . D'après les calculs précédents, on voit que

$$\begin{aligned} & \text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} - \text{cov}\{\mathbb{C}(u, v), \mathbb{C}(s, t)\} \\ &= -C_{10}(s, t)\{C(u \wedge s, v) - sC(u, v)\} \\ & \quad - C_{01}(s, t)\{C(u, v \wedge t) - tC(u, v)\} \\ & \quad - C_{10}(u, v)\{C(u \wedge s, t) - uC(s, t)\} \\ & \quad + C_{10}(u, v)C_{10}(s, t)(u \wedge s - us) \\ & \quad + C_{10}(u, v)C_{01}(s, t)\{C(u, t) - ut\} \\ & \quad - C_{01}(u, v)\{C(s, v \wedge t) - vC(s, t)\} \\ & \quad + C_{01}(u, v)C_{10}(s, t)\{C(s, v) - vs\} \\ & \quad + C_{01}(u, v)C_{01}(s, t)(v \wedge t - vt). \end{aligned}$$

En utilisant le fait que  $C(u, v) = uv$ ,  $C_{10}(u, v) = \partial C(u, v)/\partial u = v$  et  $C_{01}(u, v) = \partial C(u, v)/\partial v = u$ , on conclut que pour tous  $u, v, s, t \in [0, 1]$ ,

$$\begin{aligned} & \text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} - \text{cov}\{\mathbb{C}(u, v), \mathbb{C}(s, t)\} \\ &= -tC(u \wedge s, v) + tsuv - sC(u, v \wedge t) + stuv - vC(u \wedge s, t) + vust + vt(u \wedge s) \\ & \quad - vtus + vsut - vsut - uC(s, v \wedge t) + 2uvst - utvs + us(v \wedge t) - usvt \\ &= 2uvst - us(v \wedge t) - vt(u \wedge s) \\ &\leq 0. \end{aligned}$$

C'est cette inégalité que Genest & Segers (2010) ont généralisée à d'autres copules que l'indépendance, comme nous le verrons au chapitre suivant.

# Chapitre 4

## Covariance asymptotique

Nous avons vu au chapitre précédent comment estimer la copule  $C$  d'une loi  $H$  à marges continues au moyen de la fonction de répartition empirique  $C_n$ . Quand les marges de  $H$  sont inconnues, nous avons aussi vu qu'on peut quand même estimer  $C$  par la copule empirique  $\hat{C}_n$ . Dans le cas bivarié, Genest & Segers (2010) ont montré, sous certaines conditions, que la covariance du processus  $\hat{\mathbb{C}}_n = \sqrt{n}(\hat{C}_n - C)$  fondé sur les pseudo-observations est plus petite que la covariance du processus  $\mathbb{C}_n = \sqrt{n}(C_n - C)$  correspondant au cas où les marges sont connues. Genest & Segers (2010) ont aussi montré que ce résultat reste vrai en dimension  $d \geq 3$  pour la copule d'indépendance, mais ils ont laissé en suspens le problème de savoir si l'inégalité est valide pour des copules multivariées autres que l'indépendance.

L'objectif de ce chapitre est d'investiguer l'inégalité de Genest & Segers (2010) en dimension  $d \geq 3$ . Nous verrons que pour certains types de dépendance positive, la variance asymptotique de  $\hat{\mathbb{C}}_n$  est plus petite que celle de  $\mathbb{C}_n$  au voisinage de l'indépendance. Ces nouveaux résultats, dus à Genest et Mesfioui, ne fournissent qu'une solution partielle au problème mais des simulations présentées à la fin du chapitre suggèrent que l'inégalité désirée reste valable en dimension  $d \geq 4$  pour les familles de copules de Clayton et de Frank.

Ce chapitre est structuré comme suit. La section 4.1 rappelle quelques résultats de Genest & Segers (2010) établis dans le cas bivarié. La section 4.2 présente de nouveaux résultats théoriques qui représentent une généralisation de ceux de Genest & Segers (2010) en dimension  $d \geq 3$ ; des illustrations appuyant ces résultats y sont aussi présentées. La section 4.3 est vouée à une étude par simulation qui tend à montrer que l'inégalité recherchée est vraie pour les copules de Clayton et de Frank, peu importe le degré de dépendance positive.

## 4.1 Cas bidimensionnel

Dans cette section, nous allons rappeler quelques résultats établis par Genest & Segers (2010) dans le cas bivarié. Pour ce faire, considérons un échantillon aléatoire  $(X_1, Y_1), \dots, (X_n, Y_n)$  tiré d'une loi de fonction de répartition continue  $H$ . Supposons que les fonction de répartition marginales  $F$  et  $G$  des variables  $X$  et  $Y$  sont connues et posons  $U_i = F(X_i)$ ,  $V_i = G(Y_i)$  pour tout  $i \in \{1, \dots, n\}$ . Dans ce cas, nous pouvons estimer la copule  $C$  par la fonction de répartition empirique  $C_n$  de l'échantillon  $(U_1, V_1), \dots, (U_n, V_n)$  définie, pour tout  $(u, v) \in [0, 1]^2$ , par

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(U_i \leq u, V_i \leq v).$$

De plus, le comportement asymptotique du processus de copule empirique  $\mathbb{C}_n$ , défini par  $\mathbb{C}_n = \sqrt{n} (C_n - C)$ , est tel que précisé au chapitre 3.

Toutefois, les fonctions de répartition marginales  $F$  et  $G$  sont le plus souvent inconnues en pratique. Dans ce cas, pour estimer la copule, on commence par estimer  $F$  et  $G$  par les fonctions de répartition empiriques  $F_n$  et  $G_n$ , puis on pose  $\hat{U}_i = F_n(X_i)$  et  $\hat{V}_i = G_n(Y_i)$  pour tout  $i \in \{1, \dots, n\}$ . Ainsi, la copule est estimée par la fonction de répartition correspondant à l'échantillon des pseudo-observations  $(\hat{U}_1, \hat{V}_1), \dots, (\hat{U}_n, \hat{V}_n)$  et  $\hat{C}_n$  est définie, pour tout  $(u, v) \in [0, 1]^2$ , par

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{U}_i \leq u, \hat{V}_i \leq v).$$

De plus, le processus  $\hat{\mathbb{C}}_n = \sqrt{n} (\hat{C}_n - C)$  converge faiblement vers un processus gaussien  $\hat{\mathbb{C}}$  défini en tout  $(u, v) \in [0, 1]^2$  par

$$\hat{\mathbb{C}}(u, v) = \mathbb{C}(u, v) - C_{10}(u, v)\mathbb{C}(u, 1) - C_{01}(u, v)\mathbb{C}(1, v),$$

où  $C_{10}$  et  $C_{01}$  dénotent respectivement les dérivées partielles de  $C$  par rapport à  $u$  et  $v$ , tandis que  $\mathbb{C}$  représente le processus gaussien limite de  $\mathbb{C}_n$  dont la structure de covariance est précisée au chapitre 3.

Nous allons maintenant rappeler quelques concepts de dépendance positive utiles pour la suite du chapitre.

**Définition 4.1.1.** Une copule  $C$  est dite positivement dépendante par quadrant (PDQ) si pour tout  $(u, v) \in [0, 1]^2$ ,  $C(u, v) \geq uv$ . De même, la copule  $C$  est dite négativement dépendante par quadrant (NDQ) si pour tout  $(u, v) \in [0, 1]^2$ ,  $C(u, v) \leq uv$ .

**Définition 4.1.2.** Une copule  $C$  est dite  $LTD(V|U)$  [respectivement  $LTD(U|V)$ ] si  $C(u, v)/u$  est décroissante en  $u$  pour tout  $v \in [0, 1]$  [respectivement  $C(u, v)/v$  est décroissante en  $v$  pour tout  $u \in [0, 1]$ ]. De façon équivalente, on a

$$LTD(V|U) \Leftrightarrow \forall_{v \in (0,1)} \quad C_{10}(u, v) \leq \frac{C(u, v)}{u}$$

et

$$LTD(U|V) \Leftrightarrow \forall_{u \in (0,1)} \quad C_{01}(u, v) \leq \frac{C(u, v)}{v}.$$

Il est facile de montrer que le concept LTD est plus fort que le concept PDQ. Plusieurs des lois bivariées classiques, dont la normale, sont LTD.

Le résultat ci-dessous, établi par Genest & Segers (2010), montre que la propriété LTD implique une relation de dominance entre les fonctions de covariance asymptotique des processus limites  $\mathbb{C}$  et  $\hat{\mathbb{C}}$ .

**Proposition 4.1.1.** Soit  $C$  une copule qui est à la fois  $LTD(U|V)$  et  $LTD(V|U)$ . Supposons que les dérivées partielles  $C_{10}$  et  $C_{01}$  existent et soient continues sur  $(0, 1)^2$ . Alors, pour tous  $u, v, s, t \in [0, 1]$ , on a

$$\text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} \leq \text{cov}\{\mathbb{C}(u, v), \mathbb{C}(s, t)\}. \quad (4.1)$$

**Exemple 4.1.2.** Soit  $C$  la copule d'indépendance définie par  $C(u, v) = uv$  pour tout  $(u, v) \in [0, 1]^2$ . Une application directe de la proposition précédente montre que

$$\text{cov}\{\hat{\mathbb{C}}(u, v), \hat{\mathbb{C}}(s, t)\} - \text{cov}\{\mathbb{C}(u, v), \mathbb{C}(s, t)\} = 2uvst - us(v \wedge t) - vt(u \wedge s),$$

qui est négative pour tous  $u, v, s, t \in [0, 1]$ , comme on l'a vu dans l'exemple 3.3.2.

Dans le cas d'une dépendance négative faible, Genest & Segers (2010) ont également obtenu le résultat suivant.

**Proposition 4.1.3.** Soit  $C$  une copule NDQ telle que les dérivées partielles  $C_{10}$  et  $C_{01}$  existent et sont continues sur  $[0, 1]^2$ . Supposons que

$$C_{10}(u, v) \leq 2 \frac{C(u, v)}{u} \quad \text{et} \quad C_{01}(u, v) \leq 2 \frac{C(u, v)}{v}.$$

Alors pour tout  $(u, v) \in [0, 1]^2$  on a

$$\text{var}\{\hat{\mathbb{C}}(u, v)\} \leq \text{var}\{\mathbb{C}(u, v)\}. \quad (4.2)$$

## 4.2 Cas multidimensionnel

Dans cette section, nous allons examiner des généralisations partielles des résultats précédents récemment obtenues par Genest et Mesfioui. Soit  $C$  une copule à  $d \geq 3$  dimensions différentiable et soit  $(U_1, \dots, U_d)$  un vecteur aléatoire distribué selon  $C$ . Pour  $i, j \in \{1, \dots, d\}$  fixés et distincts, nous notons par  $C_{ij}$  la copule associée à la paire aléatoire  $(U_i, U_j)$ . Dans le contexte multivarié, on dit que la copule  $C$  est LTD si

$$\Pr \left( \bigcap_{j \neq i} \{U_j \leq u_j\} \mid U_i \leq u_i \right)$$

est une fonction décroissante en  $u_i$  pour tout  $u_j, j \neq i$  ou, d'une manière équivalente,  $C(\mathbf{u})/u_i$  est une fonction décroissante en  $u_i$  pour tout  $u_j, j \neq i$ .

Si on note  $\dot{C}_i(u_1, \dots, u_d) = \partial C(u_1, \dots, u_d) / \partial u_i$  pour tout  $i \in \{1, \dots, d\}$ , la condition ci-dessus revient à dire que pour tout  $i \in \{1, \dots, d\}$  et tous  $u_1, \dots, u_d \in (0, 1)$ , on a

$$\dot{C}_i(u_1, \dots, u_d) \leq C(u_1, \dots, u_d) / u_i.$$

Le résultat ci-dessous étend l'inégalité de Genest & Segers (2010) au cas d'une copule multivariée  $C$  caractérisée par une dépendance faible.

**Proposition 4.2.1.** *Soit  $C$  une copule à  $d \geq 2$  dimensions dont les dérivées partielles existent et sont continues. Supposons que la copule  $C$  est LTD et que pour tous  $i, j \in \{1, \dots, d\}$ , on ait*

$$\forall_{u,v \in (0,1)} \quad C_{ij}(u, v) \leq \frac{d-2}{d-1} uv + \frac{v}{d-1}. \quad (4.3)$$

Alors pour tout  $(u_1, \dots, u_d) \in [0, 1]^d$ , on a

$$\text{var}\{\hat{\mathbb{C}}(u_1, \dots, u_d)\} \leq \text{var}\{\mathbb{C}(u_1, \dots, u_d)\}. \quad (4.4)$$

**Démonstration.** On sait que pour tout  $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ , on a

$$\hat{\mathbb{C}}(\mathbf{u}) = \mathbb{C}(\mathbf{u}) - \sum_{i=1}^d \dot{C}_i(\mathbf{u}) \mathbb{C}(1, \dots, u_i, \dots, 1).$$

Ainsi,

$$\begin{aligned} \text{var}\{\hat{\mathbb{C}}(\mathbf{u})\} &= \text{var}\{\mathbb{C}(\mathbf{u})\} - 2 \sum_{i=1}^d \dot{C}_i(\mathbf{u}) \text{cov}\{\mathbb{C}(\mathbf{u}), \mathbb{C}(1, \dots, u_i, \dots, 1)\} \\ &\quad + \sum_{i=1}^d \sum_{j=1}^d \dot{C}_i(\mathbf{u}) \dot{C}_j(\mathbf{u}) \text{cov}\{\mathbb{C}(1, \dots, u_i, \dots, 1), \mathbb{C}(1, \dots, u_j, \dots, 1)\}. \end{aligned}$$



On note que  $\text{cov}\{\mathbb{C}(\mathbf{u}), \mathbb{C}(1, \dots, u_i, \dots, 1)\} = (1 - u_i)C(\mathbf{u})$  et

$$\text{cov}\{\mathbb{C}(1, \dots, u_i, \dots, 1), \mathbb{C}(1, \dots, u_j, \dots, 1)\} = \begin{cases} C_{ij}(u_i, u_j) - u_i u_j & \text{si } i \neq j; \\ u_i(1 - u_i) & \text{si } i = j. \end{cases}$$

Ainsi,

$$\text{var}\{\hat{\mathbb{C}}(\mathbf{u})\} - \text{var}\{\mathbb{C}(\mathbf{u})\}$$

$$\begin{aligned} &= \sum_{i=1}^d \dot{C}_i(\mathbf{u}) \sum_{j \neq i}^d \dot{C}_j(\mathbf{u}) \{C_{ij}(u_i, u_j) - u_i u_j\} \\ &\quad + \sum_{i=1}^d \dot{C}_i^2(\mathbf{u}) u_i(1 - u_i) - 2 \sum_{i=1}^d \dot{C}_i(\mathbf{u}) C(\mathbf{u})(1 - u_i) \\ &= \sum_{i=1}^d \dot{C}_i^2(\mathbf{u}) u_i(1 - u_i) - \sum_{i=1}^d \dot{C}_i(\mathbf{u}) C(\mathbf{u})(1 - u_i) \\ &\quad + \sum_{i=1}^d \dot{C}_i(\mathbf{u}) \sum_{j \neq i}^d \left[ \dot{C}_j(\mathbf{u}) \{C_{ij}(u_i, u_j) - u_i u_j\} - \frac{(1 - u_i)C(\mathbf{u})}{d - 1} \right]. \end{aligned}$$

Puisque  $0 \leq \dot{C}_i(\mathbf{u}) \leq C_i(\mathbf{u})/u_i$ , le premier grand terme de cette somme est négatif. En effet, on a

$$\begin{aligned} &\sum_{i=1}^d \dot{C}_i^2(\mathbf{u}) u_i(1 - u_i) - \sum_{i=1}^d \dot{C}_i(\mathbf{u}) C(\mathbf{u})(1 - u_i) \\ &\leq \sum_{i=1}^d \dot{C}_i(\mathbf{u}) C(\mathbf{u})(1 - u_i) - \sum_{i=1}^d \dot{C}_i(\mathbf{u}) C(\mathbf{u})(1 - u_i) = 0. \end{aligned}$$

La condition (4.3) entraîne aussi que le second grand terme de la somme est négatif, car

$$\begin{aligned} &\sum_{j \neq i}^d \left[ \dot{C}_j(\mathbf{u}) \{C_{ij}(u_i, u_j) - u_i u_j\} - \frac{(1 - u_i)C(\mathbf{u})}{d - 1} \right] \\ &\leq C(\mathbf{u}) \sum_{j \neq i}^d \left[ \frac{\{C_{ij}(u_i, u_j) - u_i u_j\}}{u_j} - \frac{(1 - u_i)}{d - 1} \right] \leq 0. \end{aligned}$$

On en conclut que  $\text{var}\{\hat{\mathbb{C}}(\mathbf{u})\} \leq \text{var}\{\mathbb{C}(\mathbf{u})\}$ , tel qu'annoncé.  $\square$

Remarquons que la condition (4.3) est valide pour  $d = 2$ . Notons aussi que le majorant de l'inégalité (4.3) peut s'écrire en fonction de la borne supérieure de Fréchet-Hoeffding dans le cas où la copule  $C$  est échangeable.

**Corollaire 4.2.2.** *Soit  $C$  une copule à  $d$  dimensions dont les dérivées partielles existent et sont continues. Supposons que la copule  $C$  est LTD et échangeable. Supposons en outre que pour tous  $i, j \in \{1, \dots, d\}$  et  $(u, v) \in [0, 1]^2$ ,*

$$C_{ij}(u, v) \leq \frac{d-2}{d-1} uv + \frac{1}{d-1} \min(u, v). \quad (4.5)$$

*Alors pour tout  $(u_1, \dots, u_d) \in [0, 1]^d$ , on a*

$$\text{var}\{\hat{\mathbb{C}}(u_1, \dots, u_d)\} \leq \text{var}\{\mathbb{C}(u_1, \dots, u_d)\}.$$

**Démonstration.** Puisque la copule  $C$  est échangeable, on a  $C_{ij}(u, v) = C_{ij}(v, u)$  pour tous  $u, v \in [0, 1]$ . Par conséquent, il découle de l'inégalité (4.3) que

$$\begin{aligned} C_{ij}(u, v) &\leq \min \left( \frac{d-2}{d-1} uv + \frac{v}{d-1}, \frac{d-2}{d-1} uv + \frac{u}{d-1} \right) \\ &= \frac{d-2}{d-1} uv + \frac{1}{d-1} \min(u, v). \end{aligned}$$

On peut donc conclure. □

**Exemple 4.2.3.** *On considère la copule multivariée de Farlie–Gumbel–Morgenstern (FGM). Cette copule est décrite par  $2^d - d - 1$  paramètres et définie en tout  $(u_1, \dots, u_d) \in [0, 1]^d$  par*

$$C(u_1, \dots, u_d) = \prod_{i=1}^d u_i \left\{ 1 + \sum_{k=2}^d \sum_{1 \leq s_1 < \dots < s_k \leq d} \theta_{s_1, \dots, s_k} \prod_{j=1}^k (1 - u_{s_j}) \right\}.$$

*On peut montrer que  $C(u_1, \dots, u_d)/u_i$  est décroissante sur  $u_i$  pour tout  $u_j$ ,  $i \neq j$  quand  $\theta_{s_1, \dots, s_k} \geq 0$ . Dans ce cas, la copule multivariée (FGM) est LTD. On remarque aussi que la copule bivariée  $C_{ij}$  est définie en tous  $u, v \in [0, 1]$  par*

$$C_{ij}(u, v) = uv + \theta_{ij} uv(1-u)(1-v),$$

*où  $\theta_{ij} \in [-1, 1]$ . Cette copule est LTD si  $\theta_{ij} \in [0, 1]$ . La condition (4.3) pour cette copule est équivalente à*

$$\forall u, v \in [0, 1] \quad \theta_{ij} u(1-v) \leq \frac{1}{d-1}$$

*ou encore  $\theta_{ij} \leq 1/(d-1)$ . Ainsi, la proposition 4.2.1 peut être appliquée à ce modèle si  $0 \leq \theta_{ij} \leq 1/(d-1)$  pour tous  $i, j \in \{1, \dots, d\}$ .*

On donne ci-dessous un autre exemple où la condition (4.3) est satisfaite.

**Exemple 4.2.4.** *Considérons une copule  $C$  à  $d$  dimensions dont les paires  $C_{ij}$  sont des copules de Ali-Mikhail-Haq, définies pour tous  $u, v \in [0, 1]$  par*

$$C_{ij}(u, v) = \frac{uv}{1 - \theta(1 - u)(1 - v)},$$

où  $\theta_{ij} \in (0, 1)$ . On peut vérifier que  $C_{ij}$  respecte la condition (4.3) si  $\theta_{ij} \leq 1/(d - 1)$ . Toutefois, des calculs standard montrent que

$$\frac{d - 2}{d - 1} uv + \frac{v}{d - 1} - C_{ij}(u, v) \geq 0$$

pour tous  $u, v \in [0, 1]$  si et seulement si

$$1 - (1 - v)\theta\{(d - 2)u + 1\} \geq 0 \quad (4.6)$$

pour tous  $u, v \in [0, 1]$ . Puisque la fonction  $(u, v) \mapsto 1 - (1 - v)\theta\{(d - 2)u + 1\}$  atteint son minimum en  $(u, v) = (1, 0)$ , l'inégalité (4.6) est équivalente à dire que  $1 - \theta(d - 1) \geq 0$  ou encore que  $\theta \leq 1/(d - 1)$ .

**Proposition 4.2.5.** *Soit  $C$  une copule à  $d$  dimensions différentiable. Supposons en outre que pour tout  $(u_1, \dots, u_d) \in (0, 1)^d$ , on ait*

$$\dot{C}_j(u_1, \dots, u_d) \leq \frac{2}{d} \frac{C(u_1, \dots, u_d)}{u_j}$$

Alors pour tout  $(u_1, \dots, u_d) \in [0, 1]^d$  on a

$$\text{var}\{\hat{\mathbb{C}}(u_1, \dots, u_d)\} \leq \text{var}\{\mathbb{C}(u_1, \dots, u_d)\}.$$

**Démonstration.** Supposons qu'il existe un scalaire  $\alpha \in (0, 1)$  tel que

$$\dot{C}_j(u_1, \dots, u_d) \leq \alpha C(u_1, \dots, u_d)/u_j$$

pour tout  $\mathbf{u} = (u_1, \dots, u_d) \in (0, 1)^d$ . Alors

$$\begin{aligned} & \text{var}\{\hat{\mathbb{C}}(\mathbf{u})\} - \text{var}\{\mathbb{C}(\mathbf{u})\} \\ &= \sum_{i=1}^d \dot{C}_i^2(\mathbf{u})u_i(1 - u_i) - \sum_{i=1}^d \alpha \dot{C}_i(\mathbf{u})C(\mathbf{u})(1 - u_i) \\ & \quad + \sum_{i=1}^d \dot{C}_i(\mathbf{u}) \sum_{j \neq i}^d \left[ \dot{C}_j(\mathbf{u})\{C_{ij}(u_i, u_j) - u_i u_j\} - \frac{(2 - \alpha)(1 - u_i)C(\mathbf{u})}{d - 1} \right]. \end{aligned}$$

Le premier terme de la formule est négatif car

$$\sum_{i=1}^d \dot{C}_i^2(\mathbf{u})u_i(1 - u_i) \leq \sum_{i=1}^d \alpha C(\mathbf{u})\dot{C}_i(\mathbf{u})(1 - u_i).$$

Le second terme de la formule est plus petit que

$$\sum_{i=1}^d \dot{C}_i(\mathbf{u}) C(\mathbf{u}) \sum_{j \neq i}^d \left[ \frac{\alpha \{C_{ij}(u_i, u_j) - u_i u_j\}}{u_j} - \frac{(2 - \alpha)(1 - u_i)}{d - 1} \right].$$

Choisissons maintenant la plus grande valeur possible de  $\alpha \in (0, 1)$  pour laquelle

$$\frac{\alpha \{C_{ij}(u_i, u_j) - u_i u_j\}}{u_j} - \frac{(2 - \alpha)(1 - u_i)}{d - 1} \leq 0.$$

En d'autres termes, cherchons  $\alpha \in (0, 1)$  tel que

$$C_{ij}(u_i, u_j) \leq u_i u_j + \frac{(2 - \alpha)(1 - u_i)u_j}{\alpha(d - 1)},$$

condition qui est toujours satisfaite lorsque

$$\frac{2 - \alpha}{\alpha(d - 1)} = 1$$

ou, de façon équivalente, quand  $\alpha = 2/d$ . Or c'est là le résultat annoncé.  $\square$

Notons que pour  $d = 2$ , on obtient (sous une autre forme) la même variance qu'énoncée dans la Proposition 1 de l'article de Genest & Segers (2010).

### 4.3 Étude numérique

Comme complément aux résultats de Genest et Mesfioui relatés à la section précédente, nous avons calculé numériquement la différence

$$\mathcal{D}_\theta(u_1, u_2, u_3, u_4) = \text{var}\{\hat{\mathbb{C}}_\theta(u_1, u_2, u_3, u_4)\} - \text{var}\{\mathbb{C}_\theta(u_1, u_2, u_3, u_4)\}$$

pour différents choix de  $u_1, \dots, u_4 \in (0, 1)$  et diverses valeurs du paramètre  $\theta$  pour les copules de Clayton, Frank et Gumbel-Hougaard. Dans chaque cas, nous avons généré d'une manière indépendante  $N = 1\,000\,000$  vecteurs  $(u_1, u_2, u_3, u_4)$  de  $(0, 1)^4$ . Pour chacun d'entre eux, nous avons alors calculé la différence  $\mathcal{D}_\theta(u_1, u_2, u_3, u_4)$ . Puis, nous avons ordonné ces valeurs, dénotées

$$\mathcal{D}_{\theta(1)} < \dots < \mathcal{D}_{\theta(1\,000\,000)}.$$

Un graphique des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  permet alors de voir si la valeur de  $\mathcal{D}_\theta(u_1, u_2, u_3, u_4)$  est négative ou non. Les calculs étant complexes, ils ont été réalisés au moyen du logiciel R. Le code se trouve en annexe.

### 4.3.1 Copule de Clayton

La copule de Clayton à quatre dimensions est définie en tous  $(u_1, u_2, u_3, u_4) \in (0, 1)^4$  et  $\theta > 0$  par

$$C(u_1, u_2, u_3, u_4) = (u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta} + u_4^{-\theta} - 3)^{-1/\theta},$$

de sorte que pour tous  $i, j \in \{1, \dots, 4\}$  tels que  $i \neq j$ , on trouve

$$\dot{C}_i(u_1, u_2, u_3, u_4) = u_i^{-\theta-1}(u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta} + u_4^{-\theta} - 3)^{-1-1/\theta}$$

et

$$C_{ij}(u_i, u_j) = (u_i^{-\theta} + u_j^{-\theta} - 1)^{-1/\theta}.$$

Le programme 1 rédigé en R à l'annexe permet de calculer

$$\begin{aligned} \mathcal{D}_\theta(u_1, u_2, u_3, u_4) &= \text{var}\{\hat{\mathbb{C}}(u_1, u_2, u_3, u_4)\} - \text{var}\{\mathbb{C}(u_1, u_2, u_3, u_4)\} \\ &= \mathcal{B}_1 + \mathcal{B}_2 - \mathcal{B}_3, \end{aligned}$$

où, en terme de  $\mathbf{u} = (u_1, u_2, u_3, u_4)$ , on pose

$$\begin{aligned} \mathcal{B}_1 &= \sum_{1 \leq i \neq j \leq 4} \dot{C}_i(\mathbf{u}) \dot{C}_j(\mathbf{u}) \{C_{ij}(u_i, u_j) - u_i u_j\} \\ &= 2\dot{C}_1(\mathbf{u}) \dot{C}_2(\mathbf{u}) \{C_{12}(u_1, u_2) - u_1 u_2\} + 2\dot{C}_1(\mathbf{u}) \dot{C}_3(\mathbf{u}) \{C_{13}(u_1, u_3) - u_1 u_3\} \\ &\quad + 2\dot{C}_1(\mathbf{u}) \dot{C}_4(\mathbf{u}) \{C_{14}(u_1, u_4) - u_1 u_4\} + 2\dot{C}_2(\mathbf{u}) \dot{C}_3(\mathbf{u}) \{C_{23}(u_2, u_3) - u_2 u_3\} \\ &\quad + 2\dot{C}_2(\mathbf{u}) \dot{C}_4(\mathbf{u}) \{C_{24}(u_2, u_4) - u_2 u_4\} + 2\dot{C}_3(\mathbf{u}) \dot{C}_4(\mathbf{u}) \{C_{34}(u_3, u_4) - u_3 u_4\}, \end{aligned}$$

$$\begin{aligned} \mathcal{B}_2 &= \sum_{i=1}^4 \dot{C}_i^2(\mathbf{u}) u_i (1 - u_i) \\ &= \dot{C}_1^2(\mathbf{u}) u_1 (1 - u_1) + \dot{C}_2^2(\mathbf{u}) u_2 (1 - u_2) + \dot{C}_3^2(\mathbf{u}) u_3 (1 - u_3) + \dot{C}_4^2(\mathbf{u}) u_4 (1 - u_4) \end{aligned}$$

et

$$\begin{aligned} \mathcal{B}_3 &= 2 \sum_{i=1}^4 \dot{C}_i(\mathbf{u}) C(\mathbf{u}) (1 - u_i) \\ &= 2C(\mathbf{u}) \{\dot{C}_1(\mathbf{u})(1 - u_1) + \dot{C}_2(\mathbf{u})(1 - u_2) + \dot{C}_3(\mathbf{u})(1 - u_3) + \dot{C}_4(\mathbf{u})(1 - u_4)\}. \end{aligned}$$

Les figures 4.1, 4.2 et 4.3 montrent le résultat de ces calculs lorsque  $\theta$  est respectivement égal à 1, 10 et 40. Dans tous les cas, il appert que  $\mathcal{D}_\theta(u_1, u_2, u_3, u_4) \leq 0$ , ce qui est encourageant.

FIGURE 4.1 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Clayton de paramètre  $\theta = 1$

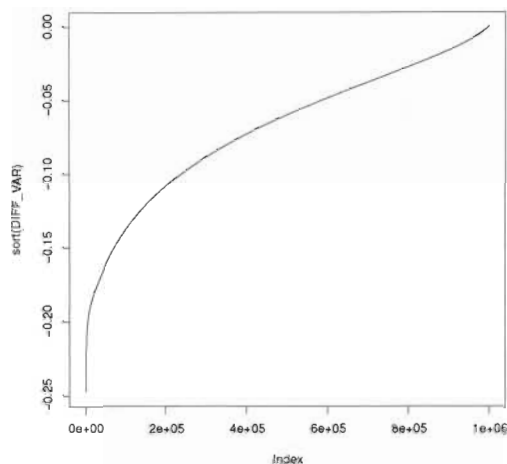


FIGURE 4.2 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Clayton de paramètre  $\theta = 10$

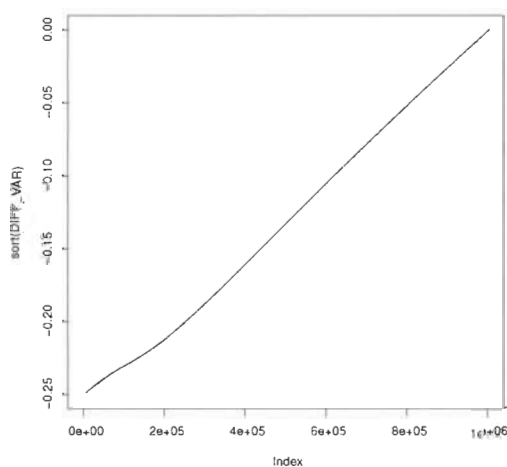
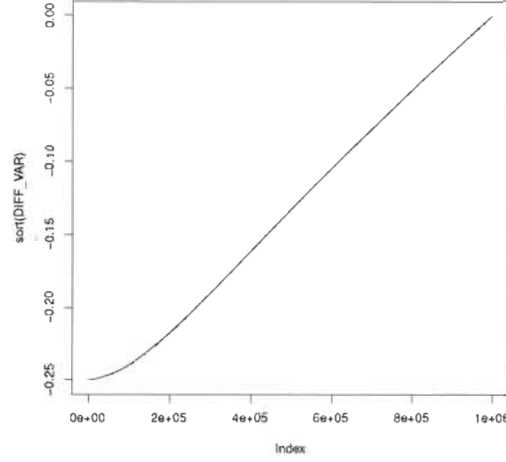


FIGURE 4.3 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Clayton de paramètre  $\theta = 40$



### 4.3.2 Copule de Frank

La copule de Frank à quatre dimensions est définie en tous  $(u_1, u_2, u_3, u_4) \in (0, 1)^4$  et  $\theta > 0$  par

$$C(u_1, u_2, u_3, u_4) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)}{(e^{-\theta} - 1)^3} \right\},$$

de sorte que pour tout  $i \in \{1, \dots, 4\}$ , on a

$$\dot{C}_i(u_1, u_2, u_3, u_4) = \frac{f_i(u_1, u_2, u_3, u_4)}{f(u_1, u_2, u_3, u_4)}$$

avec

$$f(u_1, u_2, u_3, u_4) = 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)}{(e^{-\theta} - 1)^3}$$

et  $f_i(u_1, u_2, u_3, u_4) = \partial f(u_1, u_2, u_3, u_4) / \partial u_i$ . On trouve alors

$$\dot{C}_1(u_1, u_2, u_3, u_4) = \frac{e^{-\theta u_1}(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)}{(e^{-\theta} - 1)^3 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)},$$

$$\dot{C}_2(u_1, u_2, u_3, u_4) = \frac{e^{-\theta u_2}(e^{-\theta u_1} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)}{(e^{-\theta} - 1)^3 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)},$$

$$\dot{C}_3(u_1, u_2, u_3, u_4) = \frac{e^{-\theta u_3}(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_4} - 1)}{(e^{-\theta} - 1)^3 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)}$$

et

$$\dot{C}_4(u_1, u_2, u_3, u_4) = \frac{e^{-\theta u_4}(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)}{(e^{-\theta} - 1)^3 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)(e^{-\theta u_4} - 1)},$$

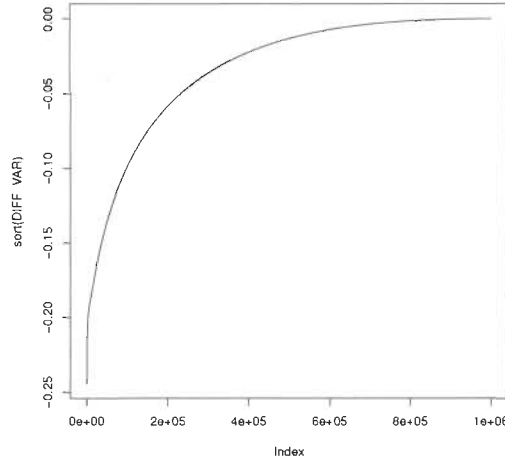
ainsi que

$$C_{ij}(u_i, u_j) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u_i} - 1)(e^{-\theta u_j} - 1)}{(e^{-\theta} - 1)} \right\}$$

pour tous  $i, j \in \{1, \dots, 4\}$  tels que  $i \neq j$ .

Le second programme en R donné en annexe permet de produire les figures 4.4, 4.5 et 4.6, lesquelles montrent le résultat des calculs lorsque  $\theta$  est respectivement égal à 1, 10 et 40. Dans tous les cas, on constate que  $\mathcal{D}_\theta(u_1, u_2, u_3, u_4) \leq 0$ , ce qui est encore une fois encourageant.

FIGURE 4.4 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Frank de paramètre  $\theta = 1$



### 4.3.3 Copule de Gumbel–Hougaard

La copule de Gumbel–Hougaard à quatre dimensions est définie en tous  $(u_1, u_2, u_3, u_4) \in (0, 1)^4$  et  $\theta > 0$  par

$$C(u_1, u_2, u_3, u_4) = \exp\{-(|\ln u_1|^\theta + |\ln u_2|^\theta + |\ln u_3|^\theta + |\ln u_4|^\theta)^{1/\theta}\}.$$



FIGURE 4.5 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Frank de paramètre  $\theta = 10$

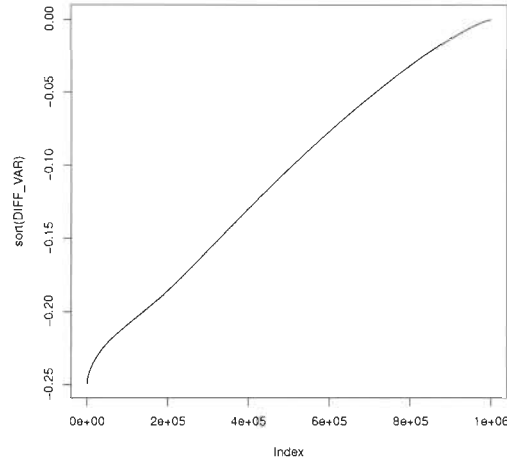
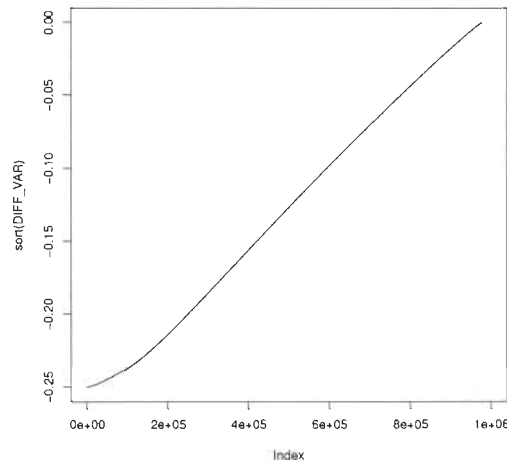


FIGURE 4.6 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Frank de paramètre  $\theta = 40$



Pour tous  $i, j \in \{1, 2, 3, 4\}$  tels que  $i \neq j$ , on trouve

$$\begin{aligned} \dot{C}_i(u_1, u_2, u_3, u_4) &= \left(\frac{1}{u_i}\right) |\ln u_i|^{\theta-1} \{(|\ln u_1|^\theta + |\ln u_2|^\theta + |\ln u_3|^\theta + |\ln u_4|^\theta)^{\frac{1-\theta}{\theta}}\} \\ &\quad \times \exp\{-(|u_1|^\theta + |\ln u_2|^\theta + |\ln u_3|^\theta + |\ln u_4|^\theta)^{1/\theta}\} \end{aligned}$$

et

$$C_{ij}(u_i, u_j) = \exp\{-(|\ln u_i|^\theta + |\ln u_j|^\theta)^{1/\theta}\}$$

pour tous  $i, j \in \{1, \dots, 4\}$  tels que  $i \neq j$ . Le programme 3 en R donné en annexe a permis de produire les figures 4.7, 4.8 et 4.9 correspondant respectivement aux cas  $\theta = 1, 10$  et  $40$ . À nouveau, on trouve  $\mathcal{D}_\theta(u_1, u_2, u_3, u_4) \leq 0$  à tout coup.

FIGURE 4.7 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Gumbel–Hougaard de paramètre  $\theta = 1$

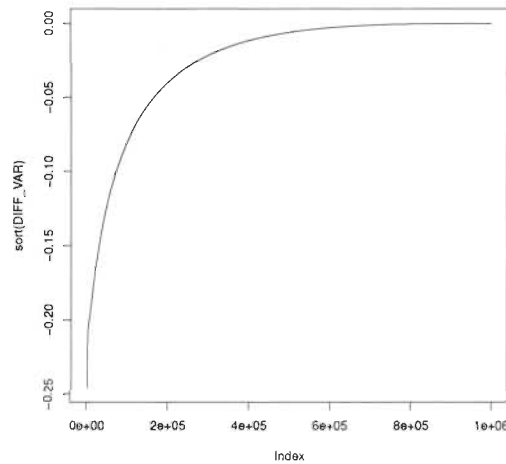


FIGURE 4.8 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Gumbel–Hougaard de paramètre  $\theta = 10$

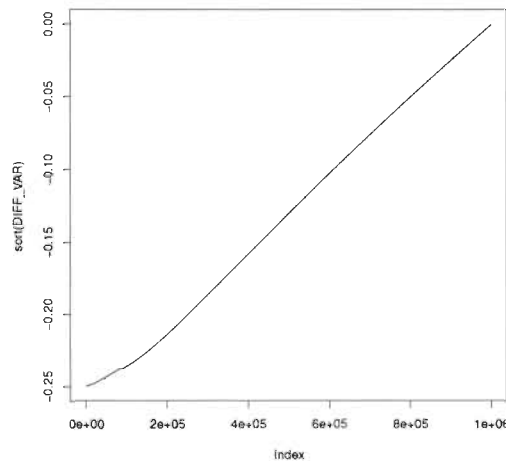
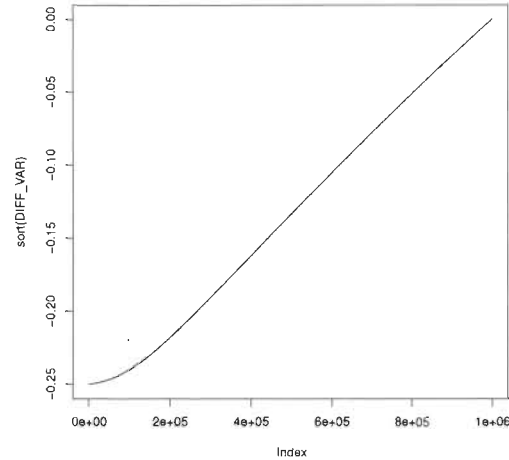


FIGURE 4.9 – Graphe des paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Gumbel–Hougaard de paramètre  $\theta = 40$



À la lumière de ces résultats numériques, on peut *conjecturer* que pour les copules archimédiennes de Clayton, de Frank et de Gumbel–Hougaard de dimension  $d = 4$ ,

$$\text{var}\{\hat{\mathbb{C}}_{\theta}(u_1, u_2, u_3, u_4)\} \leq \text{var}\{\mathbb{C}_{\theta}(u_1, u_2, u_3, u_4)\}$$

pour tous  $(u_1, u_2, u_3, u_4) \in (0, 1)^4$  et  $\theta \geq 0$ .

# Chapitre 5

## Conclusion

Genest & Segers (2010) ont montré que, sous certaines conditions, la covariance limite du processus de copule empirique  $\hat{\mathbb{C}}_n$  fondé sur les pseudo-observations est plus petite que celle du processus de copule empirique  $\mathbb{C}_n$  s'appuyant sur la connaissance des marges.

Dans ce travail, diverses généralisations de ce résultat ont été explorées. Partant de l'observation de Genest & Segers (2010) à l'effet que ce résultat reste valide en dimension arbitraire pour la copule d'indépendance, nous avons relaté et illustré des résultats partiels inédits de Genest et Mesfioui valables en dimension  $d \geq 3$ .

La méthodologie employée dérive de la démarche utilisée dans Genest & Segers (2010) dans le cas bivarié mais ne permet de démontrer que l'inégalité des variances limites des processus  $\mathbb{C}_n$  et  $\hat{\mathbb{C}}_n$  pour des copules à dépendance positive situées dans un voisinage de l'indépendance.

Toutefois, une étude numérique sur les copules de Clayton, Frank et Gumbel-Hougaard suggère que l'inégalité est vérifiée pour toutes les copules de ces trois classes en dimension  $d = 4$ . Dans des travaux ultérieurs, il serait intéressant d'explorer plus à fond la conjecture énoncée à la fin du chapitre 4.

# BIBLIOGRAPHIE

- P. Billingsley (2012). *Probability and Measure*, 4<sup>e</sup> édition. Wiley, New-York.
- M.D. Donsker (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.*, 23, 277–281.
- J.-D. Fermanian, D. Radulović & M. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10, 847–860.
- P. Gänßler & W. Stute (1987). *Seminar on Empirical Processes*. Birkhäuser, Bâle, Suisse.
- C. Genest & R.J. MacKay (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canad. J. Statist.*, 14, 145–159.
- C. Genest & J. Nešlehová (2012). Copulas and copula models. Dans *Encyclopedia of Environmetrics*, 2<sup>e</sup> édition, vol. 2. Wiley, Chichester.
- C. Genest & B. Rémillard (2004). Tests of independence and randomness based on the empirical copula process. *TEST*, 13, 335–370.
- C. Genest & J. Segers (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal.*, 101, 1837–1845.
- A.J. McNeil & J. Nešlehová (2009). Multivariate Archimedean copulas,  $d$ -monotone functions and  $\ell_1$ -norm symmetric distributions. *Ann. Statist.*, 37, 3059–3097.
- T. Mikosch (2007). Copulas : Tales and facts (suivi de commentaires). *Extremes*, 9, 1–62.
- R.B. Nelsen (2006). *An Introduction to Copulas*, 2<sup>e</sup> édition. Springer, Berlin.
- J.-F. Quessy (2014). *Notes de cours sur les copules*. Département de mathématiques et informatique, Université du Québec à Trois-Rivières, Trois-Rivières (Québec) Canada.
- L. Rüschendorf (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.*, 4, 912–923.
- J. Segers (2012). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18, 764–782.
- A. Sklar (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

# Annexe A

## Code R

**Programme 1** : Code permettant de tracer les paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Clayton de paramètre  $\theta = 10$  :

```
a<-10
N<-1000000

u1<-runif(N)
u2<-runif(N)
u3<-runif(N)
u4<-runif(N)

C<-(u1^(-a)+u2^(-a)+u3^(-a)+u4^(-a)-3)^(-1/a)

C1<-u1^(-a-1)*(u1^(-a)+u2^(-a)+u3^(-a)+u4^(-a)-3)^(-1/a-1)
C2<-u2^(-a-1)*(u1^(-a)+u2^(-a)+u3^(-a)+u4^(-a)-3)^(-1/a-1)
C3<-u3^(-a-1)*(u1^(-a)+u2^(-a)+u3^(-a)+u4^(-a)-3)^(-1/a-1)
C4<-u4^(-a-1)*(u1^(-a)+u2^(-a)+u3^(-a)+u4^(-a)-3)^(-1/a-1)

C12<-(u1^(-a)+u2^(-a)-1)^(-1/a)
C13<-(u1^(-a)+u3^(-a)-1)^(-1/a)
C14<-(u1^(-a)+u4^(-a)-1)^(-1/a)
C23<-(u2^(-a)+u3^(-a)-1)^(-1/a)
C24<-(u2^(-a)+u4^(-a)-1)^(-1/a)
C34<-(u3^(-a)+u4^(-a)-1)^(-1/a)

B1<-2*C1*C2*(C12-u1*u2)+2*C1*C3*(C13-u1*u3)+2*C1*C4*(C14-u1*u4)+
```

```
2*C2*C3*(C23-u2*u3)+2*C2*C4*(C24-u2*u4)+2*C3*C4*(C34-u3*u4)
```

```
B2<-C1^2*u1*(1-u1)+C2^2*u2*(1-u2)+C3^2*u3*(1-u3)+C4^2*u4*(1-u4)
```

```
B3<-2*C*C1*(1-u1)+2*C*C2*(1-u2)+2*C*C3*(1-u3)+2*C*C4*(1-u4)
```

```
DIFF_VAR<-B1+B2-B3
```

```
DIFF_VAR
```

```
plot(sort(DIFF_VAR), type="l")
```

**Programme 2 :** Code permettant de tracer les paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Frank de paramètre  $\theta = 10$  :

```
a<-10
```

```
N<-1000000
```

```
u1<-runif(N)
```

```
u2<-runif(N)
```

```
u3<-runif(N)
```

```
u4<-runif(N)
```

```
C<- -log(1+(exp(-a*u1)-1)*(exp(-a*u2)-1)*(exp(-a*u3)-1)*
(exp(-a*u4)-1)/(exp(-a)-1)^3)/a
```

```
C1<- exp(-a*u1)*(exp(-a*u2)-1)*(exp(-a*u3)-1)*(exp(-a*u4)-1)/
((exp(-a)-1)^3+(exp(-a*u1)-1)*(exp(-a*u2)-1)*(exp(-a*u3)-1)*(exp(-a*u4)-1))
```

```
C2<- exp(-a*u2)*(exp(-a*u1)-1)*(exp(-a*u3)-1)*(exp(-a*u4)-1)/
((exp(-a)-1)^3+(exp(-a*u1)-1)*(exp(-a*u2)-1)*(exp(-a*u3)-1)*(exp(-a*u4)-1))
```

```
C3<- exp(-a*u3)*(exp(-a*u1)-1)*(exp(-a*u2)-1)*(exp(-a*u4)-1)/
((exp(-a)-1)^3+(exp(-a*u1)-1)*(exp(-a*u2)-1)*(exp(-a*u3)-1)*(exp(-a*u4)-1))
```

```
C4<- exp(-a*u4)*(exp(-a*u1)-1)*(exp(-a*u2)-1)*(exp(-a*u3)-1)/
((exp(-a)-1)^3+(exp(-a*u1)-1)*(exp(-a*u2)-1)*(exp(-a*u3)-1)*(exp(-a*u4)-1))
```

```
C12<- -log(1+(exp(-a*u1)-1)*(exp(-a*u2)-1)/(exp(-a)-1))/a
```

```
C13<- -log(1+(exp(-a*u1)-1)*(exp(-a*u3)-1)/(exp(-a)-1))/a
```

```
C14<- -log(1+(exp(-a*u1)-1)*(exp(-a*u4)-1)/(exp(-a)-1))/a
```

```
C23<- -log(1+(exp(-a*u2)-1)*(exp(-a*u3)-1)/(exp(-a)-1))/a
```

```

C24<- -log(1+(exp(-a*u2)-1)*(exp(-a*u4)-1)/(exp(-a)-1))/a
C34<- -log(1+(exp(-a*u3)-1)*(exp(-a*u4)-1)/(exp(-a)-1))/a

B1<-2*C1*C2*(C12-u1*u2)+2*C1*C3*(C13-u1*u3)+2*C1*C4*(C14-u1*u4)+
2*C2*C3*(C23-u2*u3)+2*C2*C4*(C24-u2*u4)+2*C3*C4*(C34-u3*u4)

B2<-C1^2*u1*(1-u1)+C2^2*u2*(1-u2)+C3^2*u3*(1-u3)+C4^2*u4*(1-u4)

B3<-2*C*C1*(1-u1)+2*C*C2*(1-u2)+2*C*C3*(1-u3)+2*C*C4*(1-u4)

DIFF_VAR<-B1+B2-B3
DIFF_VAR
plot(sort(DIFF_VAR), type="l")

```

**Programme 3 :** Code permettant de tracer les paires  $(i, \mathcal{D}_{\theta(i)})$  pour  $i \in \{1, \dots, 1\,000\,000\}$  pour la copule de Gumbel–Hougaard de paramètre  $\theta = 10$  :

```

a<-10
N<-1000000

u1<-runif(N)
u2<-runif(N)
u3<-runif(N)
u4<-runif(N)

C<-exp(-(((-log(u1))^(a)+ (-log(u2))^(a)+
(-log(u3))^(a)+ (-log(u4))^(a))^(1/a))

C1<-(1/u1)*(-log(u1))^(a-1)*(((log(u1))^(a)+
(-log(u2))^(a)+ (-log(u3))^(a)+(-log(u4))^(a))^(1-a)/a))
*exp(-(((-log(u1))^(a)+ (-log(u2))^(a)+ (-log(u3))^(a)+
(-log(u4))^(a))^(1/a))

C2<-(1/u2)*(-log(u2))^(a-1)*(((log(u1))^(a)+
(-log(u2))^(a)+ (-log(u3))^(a)+(-log(u4))^(a))^(1-a)/a))
*exp(-(((-log(u1))^(a)+ (-log(u2))^(a)+ (-log(u3))^(a)+
(-log(u4))^(a))^(1/a))

C3<-(1/u3)*(-log(u3))^(a-1)*(((log(u1))^(a)+
(-log(u2))^(a)+ (-log(u3))^(a)+(-log(u4))^(a))^(1-a)/a))

```



```

*exp(-((-log(u1))^(a)+ (-log(u2))^(a)+ (-log(u3))^(a)+
(-log(u4))^(a))^(1/a))

C4<-(1/u4)*(-log(u4))^(a-1)*(((log(u1))^(a)+
(-log(u2))^(a)+ (-log(u3))^(a)+(-log(u4))^(a))^(1-a)/a))
*exp(-((-log(u1))^(a)+ (-log(u2))^(a)+ (-log(u3))^(a)+
(-log(u4))^(a))^(1/a))

C12<-exp(-((-log(u1))^(a)+ (-log(u2))^(a))^(1/a))
C13<-exp(-((-log(u1))^(a)+ (-log(u3))^(a))^(1/a))
C14<-exp(-((-log(u1))^(a)+ (-log(u4))^(a))^(1/a))
C23<-exp(-((-log(u2))^(a)+ (-log(u3))^(a))^(1/a))
C24<-exp(-((-log(u2))^(a)+ (-log(u4))^(a))^(1/a))
C34<-exp(-((-log(u3))^(a)+ (-log(u4))^(a))^(1/a))

B1<-2*C1*C2*(C12-u1*u2)+2*C1*C3*(C13-u1*u3)
+2*C1*C4*(C14-u1*u4)+2*C2*C3*(C23-u2*u3)
+2*C2*C4*(C24-u2*u4)+2*C3*C4*(C34-u3*u4)

B2<-C1^2*u1*(1-u1)+C2^2*u2*(1-u2)
+C3^2*u3*(1-u3)+C4^2*u4*(1-u4)

B3<-2*C*C1*(1-u1)+2*C*C2*(1-u2)
+2*C*C3*(1-u3)+2*C*C4*(1-u4)

DIFF_VAR<-B1+B2-B3
DIFF_VAR
plot(sort(DIFF_VAR), type="l")

```